

A. D. Aleksandrov, A. N. Kolmogorov,
M. A. Laurentiev y otros

La matemática: 2. su contenido, métodos y significado

Alianza Universidad



LA MATEMATICA: SU CONTENIDO, METODOS Y SIGNIFICADO ofrece una panorámica difícilmente mejorable en ese vasto campo temático. Publicada originalmente en la Unión Soviética, su traducción al inglés —bajo el patrocinio de la «American Mathematical Society»— ha sido saludada con grandes elogios por los críticos especializados: «una principalísima contribución a la cultura científica de nuestro siglo» (Harry Schwartz en «The New York Times Book Review»); «la mejor respuesta dada en forma de libro para la enseñanza de las matemáticas tanto en nivel elemental como en el avanzado» (Morris Kline); «una formidable y fabulosa obra maestra» («Social Science»). El tratado se mueve en dos planos: el histórico, que examina el desarrollo de la disciplina desde los orígenes hasta nuestros días, subraya la relación directísima, eclipsada en ocasiones por su alto nivel de abstracción, que las matemáticas guardan con el mundo material; y el expositivo, que logra una organización a gran escala de la matemática moderna. Los tres volúmenes (AU 68, 69 y 70) se subdividen en tres partes y diecinueve capítulos, que se ocupan de las principales ramas: análisis, geometría analítica, teoría de las ecuaciones algebraicas, ecuaciones diferenciales ordinarias, curvas y superficies, cálculo de variaciones, números primos, teoría de probabilidades, aproximación de funciones, álgebra lineal, geometrías no euclidianas, topología, análisis funcional, etcétera. P.S. ALEKSANDROV, I.M. GELFAND, A.S. KOLMOGOROV, M.A. LAURENTIEV y el resto de los diecisiete autores que han colaborado en la empresa —«los más destacados nombres de las matemáticas rusas», ha señalado «Science Fortnightly»— discutieron el plan general de la obra y las sucesivas versiones de cada capítulo, solicitando en determinados casos la opinión y colaboración de otros matemáticos soviéticos. La obra ofrece también grandes virtualidades didácticas; como ha señalado Roger S. Pinkham, «su elevada calidad expositiva y su insistencia en los conceptos más que en las demostraciones contribuirán decisivamente a elevar el nivel matemático de todos los que la estudien».

Alianza Editorial

ISBN 84-206-2069-6



Cubierta: Daniel Gil

9 788420 620695

La matemática:

su contenido,

métodos y

significado

2.

Alianza Universidad

A. D. Aleksandrov,
A. N. Kolmogorov,
M. A. Laurentiev
y otros

La matemática:
su contenido,
métodos y
significado
2.

Versión española de:
Eduardo Abad Rius

Alianza
Editorial

Título de la versión inglesa:

Mathematics: Its Content, Methods, and Meaning

Primera edición en «Alianza Universidad»: 1973

Séptima reimpresión en «Alianza Universidad»: 1994

Reservados todos los derechos. De conformidad con lo dispuesto en el art. 534-bis del Código Penal vigente, podrán ser castigados con penas de multa y privación de libertad quienes reprodujeran o plagiaran, en todo o en parte, una obra literaria, artística o científica fijada en cualquier tipo de soporte sin la preceptiva autorización.

© The Massachussets Institute of Technology

© Ed. cast.: Alianza Editorial, S. A., Madrid, 1973, 1976, 1979, 1981, 1982, 1985, 1991, 1994

Calle Juan Ignacio Luca de Tena, 15; 28027 Madrid; teléf. 741 66 00

ISBN: 84-206-2993-6 (Obra completa)

ISBN: 84-206-2069-6 (Tomo II)

Depósito legal: M. 29.918-1994

Impreso en Closas-Orcoyen, S. L. Polígono Igarsa

Paracuellos de Jarama (Madrid)

Printed in Spain

INDICE

PROLOGO A LA EDICION RUSA.....	9
--------------------------------	---

Parte 3

6. ECUACIONES DIFERENCIALES EN DERIVADAS PARCIALES S. L. Sovolev y O. A. Ladyzenskaia

§1. Introducción.....	15
§2. Las ecuaciones más simples de la física matemática.....	18
§3. Problemas iniciales y de contorno; unicidad de una solución..	29
§4. La propagación de ondas.....	40
§5. Métodos de construcción de soluciones.....	43
§6. Soluciones generalizadas.....	67

7. CURVAS Y SUPERFICIES A. D. Aleksandrov

§1. Problemas y métodos en la teoría de curvas y superficies.....	75
§2. La teoría de curvas.....	80
§3. Conceptos básicos en la teoría de superficies.....	96
§4. Geometría intrínseca y deformación de superficies.....	114
§5. Nuevos desarrollos en la teoría de curvas y superficies.....	134

8. EL CALCULO DE VARIACIONES V. I. Krilov

§1. Introducción.....	145
§2. Las ecuaciones diferenciales del cálculo de variaciones.....	151
§3. Métodos de resolución aproximada de problemas en el cálculo de variaciones.....	164

9. FUNCIONES DE UNA VARIABLE COMPLEJA

M. V. Keldysch

§1. Números complejos y funciones de una variable compleja....	167
§2. La conexión entre las funciones de una variable compleja y los problemas de física matemática.....	183
§3. La conexión entre las funciones de variable compleja y la geometría.....	195
§4. La integral de línea; fórmula de Cauchy y sus corolarios....	207
§5. Propiedades de unicidad y prolongación analítica.....	222
§6. Conclusión.....	230

Parte 4

10. NUMEROS PRIMOS

K. K. Mardzanisvili y A. B. Postnikov

§1. El estudio de la teoría de números.....	235
§2. La investigación de problemas relativos a números primos....	241
§3. Método de Chebichev.....	249
§4. Método de Vinogradov.....	256
§5. Descomposición de enteros en suma de dos cuadrados; enteros complejos.....	265

11. LA TEORIA DE PROBABILIDADES

A. N. Kolmogorov

§1. Las leyes de la probabilidad.....	269
§2. Los axiomas y fórmulas básicas de la teoría elemental de probabilidades.....	272
§3. La ley de los grandes números y teoremas del límite.....	280
§4. Observaciones adicionales sobre los conceptos básicos de la teoría de probabilidades.....	291
§5. Procesos determinados y aleatorios.....	299
§6. Procesos aleatorios de tipo Markov.....	306

12. APROXIMACION DE FUNCIONES

S. M. Nikolski

§1. Introducción.....	311
§2. Polinomios de interpolación.....	316
§3. Aproximación de integrales definidas.....	324
§4. Concepto de aproximación uniforme óptima de Chebichev....	330
§5. Los polinomios de Chebichev cuya desviación de cero es mínima.	334
§6. El teorema de Weierstrass; la aproximación óptima de una función referida a sus propiedades de diferenciabilidad.....	337
§7. Series de Fourier.....	341
§8. Aproximación en el sentido del cuadrado medio.....	349

13. METODOS DE APROXIMACION Y TECNICAS DE CALCULO

V. I. Krilov

§1. Aproximación y métodos numéricos.....	355
§2. Los medios auxiliares más simples del cálculo.....	373

PROLOGO A LA EDICION RUSA

Las matemáticas, surgidas en la Antigüedad por necesidades de la vida cotidiana, se han convertido en un inmenso sistema de variadas y extensas disciplinas. Como las demás ciencias, reflejan las leyes del mundo material que nos rodea y sirven de potente instrumento para el conocimiento y dominio de la naturaleza. Pero el alto nivel de abstracción que caracteriza a las matemáticas trae consigo el que todas sus ramas sean relativamente inaccesibles a los no especialistas. Esta cualidad abstracta de las matemáticas dio lugar, ya en la Antigüedad, a nociones idealistas sobre su independencia respecto del mundo material.

En la preparación del presente volumen, la finalidad perseguida por los autores ha sido la de familiarizar a un círculo suficientemente amplio de la intelligentsia soviética con las diversas disciplinas matemáticas, su contenido y métodos, los fundamentos sobre los que descansan y los caminos a lo largo de los cuales se han desarrollado.

El bagaje mínimo de conocimientos matemáticos que precisa el lector de este libro son las matemáticas a nivel de enseñanza media, aunque los volúmenes difieren en lo que toca a la accesibilidad del material tratado. El lector que desee iniciarse en los elementos de

las matemáticas superiores puede leer con provecho algunos de los primeros capítulos; sin embargo, para una completa comprensión de las partes subsiguientes le será necesario recurrir a los correspondientes libros de texto. El libro en su totalidad podrá ser asimilado de manera fundamental solamente por aquellos lectores que ya tengan algún conocimiento de las aplicaciones del análisis matemático, es decir, del cálculo diferencial y del cálculo integral. Para tales lectores —profesores de matemáticas, ingeniería y ciencias naturales— será particularmente importante la lectura de aquellos capítulos en los que se introducen las ramas más recientes de las matemáticas.

Naturalmente, dentro de los límites de un solo libro no ha sido posible agotar toda la riqueza ni siquiera de los resultados más fundamentales de la investigación matemática; una cierta libertad en la elección del material ha sido inevitable. Pese a ello, en líneas generales el presente libro dará una idea de la situación actual de las matemáticas, sus orígenes y su probable desarrollo futuro. Por esta razón, la obra está pensada también, hasta cierto punto, para personas ya familiarizadas con la mayoría de los temas contenidos en ella, lo cual quizá contribuya a eliminar esa estrechez de miras que se descubre ocasionalmente en algunos de nuestros jóvenes matemáticos.

Los distintos capítulos del libro están escritos por diversos autores cuyos nombres se dan en el índice; no obstante, el libro en su conjunto ha sido escrito en colaboración. El plan general, la elección del material, las sucesivas versiones de capítulos concretos, todo ello fue sometido a discusión general, introduciendo mejoras aquí y allá tras un animado cambio de opiniones. Matemáticos de distintas ciudades de la Unión Soviética tuvieron la oportunidad, mediante discusiones organizadas, de hacer numerosas y valiosas observaciones concernientes a la versión original del texto. Sus opiniones y sugerencias han sido tenidas en cuenta por los autores.

Los autores de algunos capítulos también tomaron parte directa en la redacción de la versión final de otros capítulos: la introducción del capítulo 2 fue escrita esencialmente por B. N. Delone, mientras que D. K. Fadiev desempeñó un activo papel en la preparación de los capítulos 4 y 18. Parte del trabajo fue también realizado por personas distintas de los autores de los diferentes capítulos: la sección 6 del capítulo 6 fue redactada por O. A. Ladyzenskaia, la sección 5 del capítulo 10 por A. G. Postnikov; sobre el texto del capítulo 5 trabajó O. A. Oleinik, y sobre el del capítulo 11, Ju. V. Prohorov.

Ciertas secciones de los capítulos 1, 2, 7, 16 fueron escritas por V. A. Zalgaller. La edición del texto final fue realizada por V. A. Zalgaller y V. S. Videnski con la colaboración de T. V. Rogozkinaia y A. P. Leonovaia.

La mayor parte de las ilustraciones fueron preparadas por E. P. Senkin.

Moscú
1956

COMITÉ DE EDITORES

Parte 3

Capítulo 6

ECUACIONES DIFERENCIALES EN DERIVADAS PARCIALES

§1. *Introducción*

En el estudio de los fenómenos naturales, las ecuaciones diferenciales en derivadas parciales aparecen con tanta frecuencia como las ordinarias. Por regla general esto ocurre en casos en que un suceso viene descrito por una función de varias variables.

Del estudio de la naturaleza surgió el tipo de ecuaciones en derivadas parciales, que en el momento presente está siendo investigado más a fondo y que probablemente es el más importante en la estructura general del saber humano: las ecuaciones de la física matemática.

Consideremos en primer lugar oscilaciones en un medio cualquiera. En virtud de tales oscilaciones todo punto del medio, que en equilibrio ocupa la posición (x, y, z) , se habrá desplazado al cabo de un tiempo t según el vector $u(x, y, z, t)$, que depende de la posición inicial del punto (x, y, z) y del tiempo t . En este caso el proceso en cuestión estará descrito por un campo vectorial. Pero es fácil ver que el conocimiento de este campo vectorial, esto es, el campo de desplazamiento de los puntos del medio, no es suficiente para una descripción completa de la oscilación. También es necesario conocer, por ejemplo, la densidad $\varrho(x, y, z, t)$ en cada punto del medio, la temperatura $T(x, y, z, t)$ y la tensión interna, por ejemplo, la fuerza que sobre

un volumen arbitrariamente elegido del cuerpo ejerce el resto del mismo.

Los sucesos y procesos físicos que ocurren en el espacio y el tiempo siempre consisten en variaciones, a lo largo del tiempo, de ciertas magnitudes físicas relativas a los puntos del espacio. Como hemos visto en el capítulo 2, estas cantidades pueden ser descritas por funciones de cuatro variables independientes, x, y, z y t , donde x, y y z son las coordenadas de un punto del espacio, y t es el tiempo.

Las magnitudes físicas pueden ser de diferentes clases. Algunas están completamente caracterizadas por sus valores numéricos, como la temperatura, la densidad y otras análogas, y se denominan escalares. Otras tienen dirección y son por lo tanto magnitudes vectoriales: velocidad, aceleración, intensidad de un campo eléctrico, etcétera. Las magnitudes vectoriales se pueden expresar por la longitud del vector y su dirección, pero también por sus «componentes», una vez descompuesto el vector en suma de tres vectores perpendiculares entre sí, por ejemplo, paralelos a los ejes de coordenadas.

En física matemática una magnitud escalar o un campo escalar se representa por una función de cuatro variables independientes, mientras que una magnitud vectorial definida sobre todo el espacio o, como se suele decir, un campo vectorial está descrito por tres funciones de estas variables. Podemos escribir tal magnitud en la forma

$$\mathbf{u}(x, y, z, t)$$

donde el tipo **negrita** indica que $\mathbf{u}$ es un vector, o en forma de tres funciones

$$u_x(x, y, z, t), \quad u_y(x, y, z, t), \quad u_z(x, y, z, t),$$

donde u_x, u_y y u_z son las proyecciones del vector sobre los ejes de coordenadas.

Además de las magnitudes vectoriales y escalares, aparecen en física entes aún más complicados, como es, por ejemplo, la expresión del estado de tensión de un cuerpo en un punto dado. Tales magnitudes se llaman tensores; una vez fijada la elección de los ejes de coordenadas, se pueden caracterizar por un conjunto de funciones de las cuatro variables independientes.

De esta forma, la descripción de clases muy distintas de fenómenos físicos viene normalmente dada por medio de varias funciones de varias variables. Por supuesto, tal descripción no puede ser absolutamente exacta.

Por ejemplo, cuando describimos la densidad de un medio a través de una función de cuatro variables independientes, ignoramos el hecho de que en un punto concreto podemos no tener densidad alguna. Los cuerpos que estudiamos tienen una estructura molecular, y las moléculas no están contiguas, sino que existen distancias finitas entre una y otra. Las distancias entre las moléculas son en general considerablemente más grandes que las dimensiones de las mismas. Así, pues, la densidad en cuestión es la razón de la masa contenida en un pequeño (aunque no extremadamente pequeño) volumen a este mismo volumen. La densidad en un punto podemos definirla como el límite de tales razones para volúmenes decrecientes. Una simplificación e idealización aún mayores se introducen en el concepto de temperatura de un medio. El calor en un cuerpo es debido al movimiento aleatorio de sus moléculas. La energía de las distintas moléculas es también distinta, pero si consideramos un volumen que contenga un gran número de ellas, entonces la energía media de sus movimientos aleatorios definirá lo que se denomina temperatura.

Análogamente, cuando hablamos de presión de un gas o un líquido sobre la pared de un recipiente, no pensamos en ella como si una partícula del líquido o gas estuviera realmente presionando contra la pared de dicho recipiente. En realidad, estas partículas, en su movimiento aleatorio, chocan contra la pared y rebotan en ella. Por tanto, entendemos por presión sobre la pared el efecto total de un gran número de impulsos recibidos por una parte de la misma; parte que es pequeña desde cualquier punto de vista macroscópico, pero extremadamente grande en comparación con las distancias entre las moléculas del líquido o gas. Sería fácil citar docenas de ejemplos de naturaleza similar. La mayoría de las magnitudes estudiadas en la física tienen exactamente el mismo carácter. La física matemática trabaja con magnitudes ideales, obtenidas mediante abstracción de las propiedades concretas de los entes físicos correspondientes y considerando solamente los valores medios de estas magnitudes.

Tal idealización puede parecer algo burda, pero, como veremos, es muy útil, ya que permite hacer un excelente análisis de muchas materias complicadas, en las cuales consideramos solamente los elementos esenciales y omitimos esas características que son secundarias desde nuestro punto de vista.

El objeto de la física matemática es el estudio de las relaciones que existen entre estos elementos idealizados; dichas relaciones se describen por medio de conjuntos de funciones de varias variables independientes.

§2. Las ecuaciones más simples de la física matemática

Las conexiones y relaciones elementales entre magnitudes físicas están expresadas por las leyes de la mecánica y de la física. Si bien estas relaciones son extremadamente variadas en su carácter, dan lugar a otras más complicadas que se derivan de ellas mediante argumentos matemáticos y que son aún más variadas. Las leyes de la mecánica y la física se pueden escribir en lenguaje matemático en forma de ecuaciones en derivadas parciales, o quizá ecuaciones integrales, que relacionan entre sí funciones desconocidas. Para comprender lo que esto significa, consideremos algunos ejemplos de ecuaciones de la física matemática.

Ecuaciones de la conservación de la masa y de la energía calorífica

Expresemos en forma matemática las leyes físicas básicas que gobiernan los movimientos de un medio.

1. En primer lugar expresemos la ley de conservación de la materia contenida en un volumen Ω que delimitamos mentalmente en el espacio y mantenemos fijo. Para este propósito debemos calcular la masa de la materia contenida en dicho volumen. La masa $M_{\Omega}(t)$ viene expresada por la integral

$$M_{\Omega}(t) = \iiint_{\Omega} \varrho(x, y, z, t) dx dy dz.$$

Esta masa no será, por supuesto, constante; en un proceso oscilatorio la densidad en cada punto variará, puesto que, en sus oscilaciones, las partículas de materia estarán unas veces dentro del volumen y otras fuera. La variación de la masa se puede hallar derivando respecto al tiempo, y viene dada por la integral

$$\frac{dM_{\Omega}}{dt} = \iiint_{\Omega} \frac{\partial \varrho}{\partial t} dx dy dz.$$

Esta tasa de variación de la masa contenida en el volumen se puede calcular también de otra manera. Podemos expresar la cantidad de materia que pasa cada segundo a través de la superficie S que limita nuestro volumen, afectando a la materia que sale de Ω de un signo

menos. Con este fin consideremos en la superficie S un elemento ds suficientemente pequeño para que se pueda suponer plano y para que tenga el mismo desplazamiento en todos sus puntos. Observamos el desplazamiento de los puntos de un segmento de la superficie durante el intervalo de tiempo de t a $t + dt$. Primeramente calculamos el vector

$$\mathbf{v} = \frac{d\mathbf{u}}{dt},$$

que representa la velocidad de cada partícula. En el tiempo dt las partículas que están sobre ds se mueven según el vector $\mathbf{v} dt$ y toman la posición ds_1 , mientras que la posición ds estará ocupada ahora por las partículas que anteriormente estaban en ds_2 (Fig. 1). Por tanto, durante este tiempo la columna de materia que sale del volumen Ω será la que antes estaba contenida entre ds_2 y ds . La altura de esta columna es igual a $v dt \cos(\mathbf{n}, \mathbf{v})$, donde $\mathbf{n}$ es la normal exterior a la superficie; el volumen de la pequeña columna será pues igual a

$$v \cos(\mathbf{n}, \mathbf{v}) ds dt,$$

y la masa igual a

$$\rho v \cos(\mathbf{n}, \mathbf{v}) ds dt.$$

Uniendo todos estos pequeños trozos obtenemos, para la cantidad de materia que sale del volumen durante el tiempo dt , la expresión

$$\iint_S \rho v \cos(\mathbf{n}, \mathbf{v}) ds dt.$$

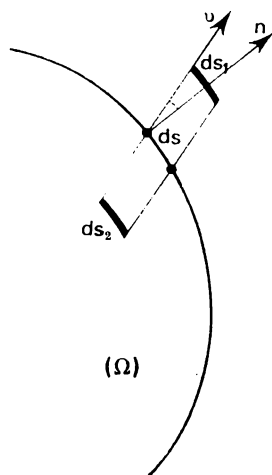


FIG. 1

En aquellos puntos en que la velocidad está dirigida hacia el interior de Ω el signo del coseno será negativo, lo que significa que en esta integral la materia que entra en Ω va afectada de un signo menos. El producto de la velocidad del movimiento del medio por su densidad se denomina flujo. El vector flujo de la masa es $\mathbf{q} = \rho \mathbf{v}$.

Para calcular la variación de flujo de materia hacia el exterior del volumen Ω basta dividir esta expresión por dt , de modo que para la variación de flujo tenemos

$$\iint_S \varrho v_n ds = \iint_S q_n ds,$$

donde

$$v_n = v \cos(\mathbf{n}, \mathbf{v}), \quad q_n = q \cos(\mathbf{n}, \mathbf{q}).$$

La componente normal del vector $\mathbf{v}$ puede reemplazarse por su expresión en términos de las componentes de los vectores $\mathbf{v}$ y $\mathbf{n}$ sobre los ejes de coordenadas. Por geometría analítica sabemos que

$$v_n = v \cos(\mathbf{n}, \mathbf{v}) = v_x \cos(\mathbf{n}, \mathbf{x}) + v_y \cos(\mathbf{n}, \mathbf{y}) + v_z \cos(\mathbf{n}, \mathbf{z});$$

así, pues, podemos volver a escribir la expresión de la variación de flujo en la forma

$$\iint_S \varrho [v_x \cos(\mathbf{n}, \mathbf{x}) + v_y \cos(\mathbf{n}, \mathbf{y}) + v_z \cos(\mathbf{n}, \mathbf{z})] ds.$$

Por la ley de conservación de la materia, estos dos métodos de calcular la variación de la cantidad de materia deben dar el mismo resultado, ya que todo cambio en la masa contenida en Ω sólo puede ser debido a la entrada o salida de masa a través de la superficie S .

Entonces, igualando la variación de la cantidad de materia contenida en el volumen a la variación de flujo de materia dentro del volumen, obtenemos

$$\begin{aligned} \iiint_{\Omega} \frac{\partial \varrho}{\partial t} dx dy dz &= \\ &= - \iint_S [\varrho v_x \cos(\mathbf{n}, \mathbf{x}) + \varrho v_y \cos(\mathbf{n}, \mathbf{y}) + \varrho v_z \cos(\mathbf{n}, \mathbf{z})] ds = \\ &= - \iint_S [q_x \cos(\mathbf{n}, \mathbf{x}) + q_y \cos(\mathbf{n}, \mathbf{y}) + q_z \cos(\mathbf{n}, \mathbf{z})] ds. \end{aligned}$$

Esta relación integral, como hemos dicho, es válida para todo volumen Ω . Se denomina «ecuación de continuidad».

La integral del segundo miembro de la última ecuación se puede transformar en una integral de volumen mediante la fórmula de Ostrogradski. Esta fórmula, deducida en el capítulo 2, nos da

$$\begin{aligned} \iint_S [\rho v_x \cos(\mathbf{n}, x) + \rho v_y \cos(\mathbf{n}, y) + \rho v_z \cos(\mathbf{n}, z)] ds = \\ = \iiint_\Omega \left[\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} \right] d\Omega. \end{aligned}$$

De aquí se deduce que

$$\iiint_\Omega \left[\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} \right] d\Omega = 0.$$

De esta forma obtenemos el siguiente resultado: la integral de la función

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} \quad \text{ó} \quad \frac{\partial \rho}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z}$$

sobre cualquier volumen Ω es cero. Pero esto sólo es posible si la función es idénticamente nula. De este modo obtenemos la ecuación de continuidad en forma diferencial

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} = 0 \quad (1)$$

La ecuación (1) es un típico ejemplo de la formulación de una ley física en el lenguaje de las ecuaciones en derivadas parciales.

2. Consideremos otro problema análogo: el problema de la conducción del calor.

En un medio cuyas partículas están en movimiento, el calor fluye de unos puntos a otros. Este flujo de calor tiene lugar a través de cada elemento de superficie ds del medio considerado. Es posible probar que el proceso se puede describir numéricamente por medio de una única magnitud vectorial, el vector conducción de calor, que se denota por τ . Entonces la cantidad de calor que fluye por segundo a través de un elemento de área ds se podrá expresar por $\tau_n ds$, de la misma

forma que $q_n ds$ expresaba la cantidad de materia que pasaba por segundo a través de un área ds . En lugar del flujo de líquido $q = \varrho v$ tenemos el vector flujo de calor τ .

Análogamente a como hemos obtenido la ecuación de continuidad, que expresa, para el movimiento de un líquido, la ley de conservación de la masa, podemos deducir una nueva ecuación en derivadas parciales que exprese la ley de conservación de la energía; veámoslo.

La densidad de energía calorífica Q en un punto dado puede expresarse mediante la fórmula

$$Q = CT,$$

donde C es la capacidad calorífica y T la temperatura.

Aquí es fácil obtener la ecuación

$$C \frac{\partial T}{\partial t} + \frac{\partial \tau_x}{\partial x} + \frac{\partial \tau_y}{\partial y} + \frac{\partial \tau_z}{\partial z} = 0. \quad (2)$$

La deducción de esta ecuación es idéntica a la de la ecuación de continuidad, sustituyendo «densidad» por «densidad de energía calorífica» y flujo de masa por flujo de calor. Aquí hemos supuesto que la energía calorífica en el medio nunca aumenta. Pero, si existe una fuente de calor en él, la ecuación (2) del equilibrio de energía calorífica debe ser modificada. Si q es la densidad de productividad de la fuente, es decir, la cantidad de energía calorífica producida por unidad de volumen en un segundo, entonces la ecuación de conservación de la energía calorífica adopta la siguiente forma, algo más complicada:

$$C \frac{\partial T}{\partial t} + \frac{\partial \tau_x}{\partial x} + \frac{\partial \tau_y}{\partial y} + \frac{\partial \tau_z}{\partial z} = q. \quad (3)$$

3. Aun otra ecuación del mismo tipo que la de continuidad se puede obtener derivando (1) respecto al tiempo. Hagámoslo para la ecuación de pequeñas oscilaciones de un gas próximo a una posición de equilibrio. Supondremos que para tales oscilaciones los cambios de densidad no son grandes y que las cantidades $\partial \varrho / \partial x$, $\partial \varrho / \partial y$, $\partial \varrho / \partial z$ y $\partial \varrho / \partial t$ son suficientemente pequeñas para que sus productos por v_x , v_y y v_z puedan despreciarse. Entonces

$$\frac{\partial \varrho}{\partial t} + \varrho \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) = 0.$$

Derivando esta ecuación con respecto al tiempo y despreciando los productos de $\partial \rho / \partial t$ con $\partial v_x / \partial x$, $\partial v_y / \partial y$ y $\partial v_z / \partial z$, obtenemos

$$\frac{\partial^2 \rho}{\partial t^2} + \rho \left[\frac{\partial \left(\frac{dv_x}{dt} \right)}{\partial x} + \frac{\partial \left(\frac{dv_y}{dt} \right)}{\partial y} + \frac{\partial \left(\frac{dv_z}{dt} \right)}{\partial z} \right] = 0. \quad (4)$$

Ecuación del movimiento

1. Un importante ejemplo de la expresión de una ley física mediante una ecuación diferencial se da en las ecuaciones de equilibrio o de movimiento de un medio. Sea un medio formado por partículas materiales moviéndose con distintas velocidades. Como en el primer ejemplo, delimitemos mentalmente en el espacio un volumen Ω encerrado por la superficie S y lleno de partículas de materia del medio, y escribamos la segunda ley de Newton para las partículas de este volumen. Dicha ley establece que la variación de la cantidad de movimiento en el volumen, sumada para todas las partículas, es igual a la suma de todas las fuerzas que actúan sobre el volumen. La cantidad de movimiento, como se sabe por la mecánica, está representada por el vector

$$\mathbf{P} = \iiint_{\Omega} \rho \mathbf{v} d\Omega.$$

Las partículas que ocupan un pequeño volumen $d\Omega$ con densidad ρ llenarán, al cabo de un tiempo Δt , un nuevo volumen $d\Omega'$ con densidad ρ' , si bien la masa será la misma

$$\rho' d\Omega' = \rho d\Omega.$$

Si la velocidad $\mathbf{v}$ se transforma durante ese tiempo en una nueva velocidad $\mathbf{v}'$, es decir, varía en el incremento $\Delta \mathbf{v} = \mathbf{v}' - \mathbf{v}$, el correspondiente incremento de la cantidad de movimiento será

$$\rho' \mathbf{v}' d\Omega' - \rho \mathbf{v} d\Omega = \rho \mathbf{v}' d\Omega - \rho \mathbf{v} d\Omega = \rho \Delta \mathbf{v} d\Omega,$$

o en la unidad de tiempo:

$$\rho \frac{\Delta \mathbf{v}}{\Delta t} d\Omega \approx \rho \frac{d\mathbf{v}}{dt} d\Omega.$$

Sumando sobre todas las partículas del volumen Ω , obtenemos que el ritmo de variación de la cantidad de movimiento es igual a

$$\iiint_{\Omega} \rho \frac{d\mathbf{v}}{dt} d\Omega,$$

o, lo que es lo mismo,

$$\iiint_{\Omega} \rho \frac{dv_x}{dt} d\Omega, \quad \iiint_{\Omega} \rho \frac{dv_y}{dt} d\Omega, \quad \iiint_{\Omega} \rho \frac{dv_z}{dt} d\Omega.$$

[Aquí las derivadas dv_x/dt , dv_y/dt y dv_z/dt denotan la variación de las componentes de $\mathbf{v}$, no en un punto dado del espacio, sino para una partícula dada. Este es el significado de utilizar la notación d/dt en vez de $\partial/\partial t$. Como es bien sabido, $d/dt = \partial/\partial t + v_x(\partial/\partial x) + v_y(\partial/\partial y) + v_z(\partial/\partial z)$.]

Las fuerzas que actúan sobre el volumen pueden ser de dos clases: fuerzas de volumen que actúan sobre cada partícula del cuerpo y fuerzas de superficie o tensiones sobre la superficie S que limita el volumen. Aquéllas son fuerzas de largo alcance, las últimas de corto alcance.

Para ilustrar estas consideraciones, supongamos que el medio en cuestión es un fluido. Las fuerzas de superficie que actúan sobre un elemento de área ds tendrán en este caso el valor $p ds$, donde p es la presión sobre el fluido, y se ejercerán en dirección opuesta a la de la normal exterior.

Si denotamos por $\mathbf{n}$ el vector unitario en la dirección de la normal a la superficie, entonces la fuerza total que actúa sobre la sección ds será igual a

$$-\mathbf{p}\mathbf{n} ds.$$

Si denotamos por $\mathbf{F}$ el vector de las fuerzas exteriores que actúan sobre una unidad de volumen, nuestra ecuación toma la forma

$$\iiint_{\Omega} \rho \frac{d\mathbf{v}}{dt} d\Omega = \iiint_{\Omega} \mathbf{F} d\Omega - \int_S \mathbf{p}\mathbf{n} ds.$$

Esta es la ecuación del movimiento en forma integral. Como la de continuidad, también se puede transformar en forma diferencial.

Obtenemos el sistema

$$\rho \frac{dv_x}{dt} + \frac{\partial p}{\partial x} = F_x, \rho \frac{dv_y}{dt} + \frac{\partial p}{\partial y} = F_y, \rho \frac{dv_z}{dt} + \frac{\partial p}{\partial z} = F_z. \quad (5)$$

Este sistema es la forma diferencial de la segunda ley de Newton.

2. Otro ejemplo característico de la aplicación de las leyes de la mecánica en forma diferencial es la ecuación de una cuerda vibrante. Una cuerda es un trozo de material elástico, alargado, muy delgado, que es flexible debido a su extrema delgadez, y que normalmente se mantiene tensado. Si imaginamos la cuerda dividida por el punto x en dos partes, sobre cada una de ellas se ejerce una fuerza igual a la tensión en la dirección de la tangente a la curva formada por la cuerda.

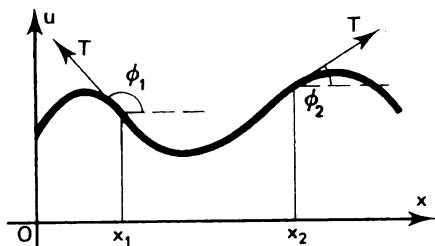


FIG. 2

Examinemos un corto

segmento de la cuerda. Denotemos por $u(x, t)$ el desplazamiento de un punto de la misma desde su posición de equilibrio. Supongamos que la cuerda oscila en un plano y que tal oscilación consiste en desplazamientos perpendiculares al eje Ox . Representemos el desplazamiento $u(x, t)$ gráficamente en un instante dado (Fig. 2), y estudiemos el comportamiento del segmento de cuerda comprendido entre los puntos x_1 y x_2 . En estos puntos actúan dos fuerzas iguales a la tensión T en la dirección de la correspondiente tangente a $u(x, t)$.

Si el segmento es curvo, la resultante de estas dos fuerzas no será nula. Esta resultante, por las leyes de la mecánica, debe ser igual a la variación de la cantidad de movimiento del segmento.

Sea ρ la masa contenida en cada centímetro de longitud de la cuerda. Entonces la variación de la cantidad de movimiento será

$$\rho \int_{x_1}^{x_2} \frac{\partial^2 u}{\partial t^2} dx.$$

Si el ángulo formado por la tangente a la cuerda y el eje Ox se denota por ϕ , tendremos

$$T \sin \phi_2 - T \sin \phi_1 = \rho \int_{x_1}^{x_2} \frac{\partial^2 u}{\partial t^2} dx.$$

Esta es la ecuación usual que expresa la segunda ley de la mecánica en forma integral. Es fácil transformarla en forma diferencial. Es evidente que:

$$\varrho \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x} (T \operatorname{sen} \phi).$$

Recurriendo a teoremas bien conocidos del cálculo diferencial, es fácil relacionar $T \operatorname{sen} \phi$ con la función desconocida u . Obtenemos

$$\operatorname{tg} \phi = \frac{\partial u}{\partial x}, \operatorname{sen} \phi = \frac{\operatorname{tg} \phi}{\sqrt{1 + \operatorname{tg}^2 \phi}} = \frac{\partial u / \partial x}{\sqrt{1 + (\partial u / \partial x)^2}},$$

y bajo la hipótesis de que $(\partial u / \partial x)^2$ es pequeño, tenemos

$$\operatorname{sen} \phi \approx \frac{\partial u}{\partial x}.$$

Entonces

$$T \frac{\partial^2 u}{\partial x^2} = \varrho \frac{\partial^2 u}{\partial t^2}. \quad (6)$$

Esta última es la *ecuación de la cuerda vibrante* en forma diferencial.

Formas básicas de las ecuaciones de la física matemática

Como ya mencionamos anteriormente, las distintas ecuaciones en derivadas parciales que describen fenómenos físicos forman normalmente un sistema de ecuaciones en varias variables desconocidas. Pero en la gran mayoría de los casos es posible reemplazar este sistema por una sola ecuación, como se puede probar fácilmente con ejemplos muy sencillos.

Por ejemplo, volvamos a las ecuaciones del movimiento consideradas en la sección anterior y tratemos de resolverlas junto con la ecuación de continuidad. Los métodos efectivos de resolución los consideraremos más adelante.

1. Comencemos con la ecuación del flujo uniforme de un fluido ideal.

Los movimientos posibles de un fluido se pueden dividir en rotacionales e irrotacionales, estos últimos también llamados *potenciales*. Aunque los irrotacionales son sólo casos especiales del movimiento

y, por lo general, el movimiento de un líquido o gas es siempre más o menos rotacional, sin embargo la experiencia prueba que en muchos casos el movimiento es prácticamente irrotacional. Por otra parte, se puede demostrar mediante consideraciones teóricas que en un fluido con viscosidad nula un movimiento que es inicialmente irrotacional persiste en este estado.

Para un movimiento potencial, existe una función escalar $U(x, y, z, t)$, llamada *potencial de velocidades*, tal que el vector velocidad viene expresado en términos de esta función mediante las fórmulas

$$v_x = \frac{\partial U}{\partial x}, \quad v_y = \frac{\partial U}{\partial y}, \quad v_z = \frac{\partial U}{\partial z}.$$

En todos los casos estudiados hasta ahora hemos tratado con sistemas de cuatro ecuaciones con cuatro funciones desconocidas o, dicho con otras palabras, con una ecuación escalar y una ecuación vectorial que contenían un campo escalar y un campo vectorial desconocidos. Normalmente estas ecuaciones se pueden combinar en una sola con una función desconocida, aunque esta ecuación será de segundo orden. Hagámoslo, empezando por el caso más simple.

Para un movimiento potencial de un fluido incompresible, para el que $\partial \varrho / \partial t = 0$, tenemos dos sistemas de ecuaciones: la ecuación de continuidad

$$\varrho \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) = 0$$

y las ecuaciones del movimiento potencial

$$v_x = \frac{\partial U}{\partial x}, \quad v_y = \frac{\partial U}{\partial y}, \quad v_z = \frac{\partial U}{\partial z}.$$

Sustituyendo en la primera los valores de la velocidad que tenemos en la segunda, queda

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0. \quad (7)$$

2. El campo vectorial de «flujo de calor» puede también expresarse, mediante ecuaciones diferenciales, en términos de una magnitud escalar, la temperatura. Es bien sabido que el calor «fluye» de los

cuerpos calientes a los fríos. Así, pues, el vector flujo de calor es de dirección opuesta al llamado vector gradiente de temperatura. También es natural suponer, como justifica la experiencia, que en una primera aproximación la longitud de dicho vector es directamente proporcional al gradiente de temperatura.

Las componentes del gradiente de temperatura son

$$\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}.$$

Tomando k como coeficiente de proporcionalidad, obtenemos tres ecuaciones

$$\tau_x = -k \frac{\partial T}{\partial x}, \tau_y = -k \frac{\partial T}{\partial y}, \tau_z = -k \frac{\partial T}{\partial z}$$

Estas ecuaciones hay que resolverlas junto con la ecuación de conservación de la energía calorífica

$$C \frac{\partial T}{\partial t} + \frac{\partial \tau_x}{\partial x} + \frac{\partial \tau_y}{\partial y} + \frac{\partial \tau_z}{\partial z} = q.$$

Sustituyendo τ_x , τ_y y τ_z por sus valores en función de T , obtenemos

$$C \frac{\partial T}{\partial t} = k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + q. \quad (8)$$

3. Finalmente, para pequeñas vibraciones en un medio gaseoso, por ejemplo las vibraciones sonoras, la ecuación

$$\frac{\partial^2 \varrho}{\partial t^2} + \varrho \frac{\partial}{\partial x} \left(\frac{dv_x}{dt} \right) + \varrho \frac{\partial}{\partial y} \left(\frac{dv_y}{dt} \right) + \varrho \frac{\partial}{\partial z} \left(\frac{dv_z}{dt} \right) = 0$$

y las ecuaciones de la dinámica (5) dan

$$\varrho \frac{dv_x}{dt} + \frac{\partial p}{\partial x} = F_x, \quad \varrho \frac{dv_y}{dt} + \frac{\partial p}{\partial y} = F_y, \quad \varrho \frac{dv_z}{dt} + \frac{\partial p}{\partial z} = F_z,$$

y, suponiendo la ausencia de fuerzas externas ($F_x = F_y = F_z = 0$), obtenemos

$$\frac{\partial^2 p}{\partial t^2} = a^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right) \quad (9)$$

(para obtener esta ecuación basta sustituir la expresión de las aceleraciones en la ecuación de continuidad y eliminar la densidad utilizando la ley de Boyle-Mariotte: $p = a^2 \rho$).

Las ecuaciones (7), (8) y (9) son típicas de muchos problemas de física matemática aparte de los considerados aquí. El hecho de que hayan sido estudiados con detalle nos permite tener un conocimiento de muchas situaciones físicas.

§3. *Problemas iniciales y de contorno; unicidad de una solución*

Tanto en las ecuaciones diferenciales en derivadas parciales como en las ordinarias ocurre, salvo raras excepciones, que toda ecuación tiene infinitas soluciones particulares. Así, pues, para resolver un problema físico concreto, es decir, para calcular una función desconocida que verifique una ecuación dada, debemos saber cómo elegir, en el conjunto infinito de soluciones, la que nos interesa. Para este propósito suele ser necesario conocer no sólo la ecuación, sino también un cierto número de condiciones suplementarias. Como vimos anteriormente, las ecuaciones en derivadas parciales expresan las leyes elementales de la mecánica o la física referentes a pequeñas partículas situadas en un medio. Pero, si queremos predecir el curso de un proceso, no basta con conocer las leyes de la mecánica. Por ejemplo, para predecir el movimiento de los astros es preciso conocer no sólo la formulación general de las leyes de Newton, sino también —suponiendo que las masas de estos cuerpos son conocidas— el estado inicial del sistema, es decir, la posición de los astros y sus velocidades en algún instante inicial. Condiciones suplementarias de esta clase se encuentran siempre en la resolución de problemas de física matemática.

Así, pues, los problemas de física matemática consisten en el cálculo de soluciones de ecuaciones en derivadas parciales que satisfagan ciertas condiciones suplementarias.

Las ecuaciones (7), (8) y (9) difieren en su estructura; de ahí que también sean diferentes los problemas físicos que pueden ser resueltos por medio de estas ecuaciones.

Las ecuaciones de Laplace y Poisson; funciones armónicas y unicidad de solución de los problemas de contorno relativos a ellas

Analicemos estos problemas con un poco más de detalle. Comencemos con las ecuaciones de Laplace y Poisson. La *ecuación de Poisson* es¹

$$\Delta u = -4\pi\rho,$$

donde ρ es usualmente la densidad. En particular, ρ puede ser nula. Para $\rho \equiv 0$ obtenemos la *ecuación de Laplace*

$$\Delta u = 0.$$

No es difícil ver que la diferencia entre dos soluciones particulares u_1 y u_2 de la ecuación de Poisson es una función que verifica la ecuación de Laplace, o, dicho con otras palabras, que es una *función armónica*. El espacio total de soluciones de la ecuación de Poisson se reduce, por tanto, al espacio de funciones armónicas.

Una vez obtenida una solución particular u_0 de la ecuación de Poisson, si definimos una nueva función desconocida w por

$$u = u_0 + w,$$

vemos que w debe verificar la ecuación de Laplace; y exactamente de la misma manera determinamos las correspondientes condiciones de contorno para w . Así, pues, es particularmente importante estudiar los problemas de contorno para la ecuación de Laplace.

El caso más frecuente en los problemas matemáticos es que el planteamiento del problema para una ecuación de la física matemática venga inmediatamente sugerido por la situación práctica. Las condiciones suplementarias que aparecen en la solución de la ecuación de Laplace provienen del enunciado físico del problema.

Consideremos, por ejemplo, el establecimiento de una temperatura estacionaria en un medio, es decir, la propagación del calor en un medio en que las fuentes de calor son constantes y están situadas dentro o fuera del medio. En estas condiciones, con el transcurso del tiempo la temperatura alcanzada en un punto de dicho medio será independiente del tiempo. Así, pues, para hallar la temperatura T en cada punto debemos calcular aquella solución de la ecuación

$$\frac{\partial T}{\partial t} = \Delta T + q,$$

¹ El símbolo Δu es una abreviatura de la expresión $\partial^2 u / \partial x^2 + \partial^2 u / \partial y^2 + \partial^2 u / \partial z^2$ y se denomina *laplaciana* de la función u .

(donde q es la densidad de fuentes de distribución de calor) que es independiente de t . Obtenemos

$$\Delta T + q = 0.$$

Por tanto, la temperatura en nuestro medio satisface la ecuación de Poisson. Si la densidad q de fuentes de calor es cero, entonces la ecuación de Poisson se convierte en la de Laplace.

Para calcular la temperatura en el interior del medio es necesario, por simples consideraciones físicas, conocer también lo que ocurre en la frontera del mismo.

Evidentemente, las leyes físicas que consideramos anteriormente para puntos interiores de un cuerpo necesitan formulación bien distinta para puntos del contorno.

En el problema de establecer un estado estacionario de temperatura podemos imponer, o la distribución de la misma en el contorno, o la variación de flujo de calor a través de una unidad de área de la superficie, o, finalmente, una ley que relacione la temperatura con el flujo de calor.

Considerando la temperatura en un volumen Ω limitado por la superficie S , podemos escribir estas tres condiciones así:

$$T|_S = \phi(Q), \quad (10)$$

o bien

$$\frac{\partial T}{\partial n}|_S = \psi(Q), \quad (10')$$

o, finalmente, en el caso más general

$$\alpha \frac{\partial T}{\partial n}|_S + \beta T|_S = \chi(Q), \quad (10'')$$

donde Q denota un punto arbitrario de la superficie S . Las condiciones de la forma (10) se denominan *condiciones de contorno*. El estudio de las ecuaciones de Laplace y Poisson bajo condiciones de contorno de uno de estos tipos demostrará que, por regla general, la solución está unívocamente determinada.

Por tanto, en nuestra búsqueda de una solución de la ecuación de Laplace o de la de Poisson, será normalmente necesario y suficiente

especificar una función arbitraria sobre la frontera del dominio². Examinemos la ecuación de Laplace con un poco más detalle. Probaremos que una función armónica u , es decir, una función que verifica la ecuación de Laplace, está completamente determinada si conocemos sus valores sobre la frontera del dominio.

En primer lugar, sentemos el hecho de que una función armónica no puede tomar en el interior del dominio valores mayores que el máximo valor sobre el contorno. Más concretamente, el máximo absoluto, así como el mínimo absoluto, de una función armónica se alcanzan en la frontera del dominio.

De aquí se deducirá que, si una función armónica es constante sobre el contorno de un dominio Ω , entonces en su interior también valdrá la misma constante. Ya que si el máximo y el mínimo de una función coinciden, entonces dicha función será en todo punto igual a este valor común.

Establezcamos ahora el hecho de que el máximo y el mínimo absolutos de una función armónica no pueden alcanzarse en el interior del dominio. En primer lugar, observemos que si la laplaciana Δu de la función $u(x, y, z)$ es positiva en todo el dominio, la función no puede tener un máximo en el interior, y si es negativa, entonces no puede tener un mínimo en el interior. En un punto donde la función alcance un máximo absoluto debe tener también un máximo como función de cada una de las variables por separado para valores fijos de las demás. Así, pues, se deduce que todas las derivadas parciales de segundo orden respecto a cada una de las variables deben ser no positivas. Esto significa que su suma será no positiva, mientras que la laplaciana es positiva, lo cual es imposible. Análogamente se puede probar que si la función tiene un mínimo en algún punto interior, entonces la laplaciana no puede ser negativa en él. Esto significa que si la laplaciana es negativa en todo el dominio, la función no puede tener un mínimo en él.

Una función armónica se puede siempre transformar mediante una cantidad arbitrariamente pequeña de tal manera que tenga una laplaciana positiva o negativa; para ello basta con sumar la cantidad

$$\pm \eta r^2 = \pm \eta(x^2 + y^2 + z^2),$$

donde η es una constante arbitrariamente pequeña.

² La expresión «función arbitraria» significa aquí y en lo que sigue que no se impone a las funciones ninguna condición especial, salvo ciertos requisitos de regularidad.

La adición de una cantidad suficientemente pequeña no puede alterar la propiedad de que la función tenga un máximo o un mínimo absolutos dentro del dominio. Si una función armónica tiene un máximo en el interior del dominio, entonces añadiéndole $+\eta r^2$ obtendríamos una función con laplaciana positiva que, como probamos antes, no podría tener máximo en el interior. Esto significa que una función armónica no puede tener un máximo absoluto interior al dominio. Análogamente se puede probar que tampoco puede tener un mínimo en el interior.

Este teorema tiene un importante corolario. Dos funciones armónicas que coincidan sobre la frontera de un dominio deben coincidir en el interior. Ya que entonces la diferencia de estas funciones (que será una función armónica) se anula sobre el contorno y por tanto es igual a cero en todo punto del interior.

Vemos así que los valores de una función armónica sobre el contorno determinan completamente la función. Se puede probar (aunque no podemos hacerlo aquí con detalle) que para valores arbitrariamente prefijados sobre el contorno se puede hallar siempre una función armónica que tome dichos valores.

Algo más complicado es probar que la temperatura estacionaria establecida en un cuerpo viene completamente determinada si conocemos la variación de flujo de calor a través de cada elemento de superficie del cuerpo o una ley que relacione el flujo de calor con la temperatura. Volveremos a algunos aspectos de esta cuestión cuando discutamos los métodos de resolución de problemas de la física matemática.

El problema de contorno para la ecuación del calor

Una situación completamente diferente la tenemos en el problema de la ecuación del calor en el caso no estacionario. Físicamente está claro que los valores de la temperatura sobre el contorno o la variación de flujo de calor a través de la frontera no son suficientes para definir la unidad de la solución del problema. Pero, si además conocemos la distribución de la temperatura en un instante inicial, entonces el problema sí está unívocamente determinado. Por tanto, para determinar la solución de la ecuación de conducción de calor (8) es normalmente necesario y suficiente designar una función arbitraria $T_0(x, y, z)$ que describa la distribución inicial de temperatura y también una función arbitraria sobre la frontera del dominio. Como antes, esta función puede ser, o la temperatura sobre la superficie del cuerpo, o la variación de flujo de calor a través de cada elemento de superficie, o una ley que relacione el flujo de calor con la temperatura.

De esta manera, el problema puede enunciarse como sigue. Buscamos una solución de la ecuación (8) bajo la condición

$$T|_{t=0} = T_0(x, y, z) \quad (11)$$

y una de las tres condiciones siguientes:

$$T|_S = \phi(Q), \quad (12)$$

$$\left. \frac{\partial T}{\partial n} \right|_S = \psi(Q), \quad (12')$$

$$\alpha \left. \frac{\partial T}{\partial n} \right|_S + \beta T|_S = \chi(Q), \quad (12'')$$

donde Q es un punto cualquiera de la superficie S .

La condición (11) se denomina una *condición inicial*, mientras que las (12) son *condiciones de contorno*.

No probaremos con detalle que todo problema de este tipo tiene una única solución; sólo estableceremos este hecho para el primero de estos problemas; además, consideraremos solamente el caso en que no existen fuentes de calor en el interior del medio. Demostraremos que la ecuación

$$\Delta T = \frac{1}{a^2} \frac{\partial T}{\partial t}$$

bajo las condiciones

$$T|_{t=0} = T_0(x, y, z),$$

$$T|_S = \phi(Q)$$

solamente puede tener una solución.

La demostración de este enunciado es muy similar a la demostración de la unicidad de la solución de la ecuación de Laplace. Probamos en primer lugar que si

$$\Delta T - \frac{1}{a^2} \frac{\partial T}{\partial t} < 0,$$

entonces T , como función de cuatro variables, x, y, z y t ($0 \leq t \leq t_0$), alcanza un mínimo sobre la frontera del dominio Ω o en el interior, si bien en este último caso necesariamente ha de ser en el instante inicial, $t = 0$.

Si no fuera así, el mínimo lo alcanzaría en un punto interior.

En este punto las derivadas primeras, incluida $\partial T/\partial t$, serían iguales a cero, y si este mínimo se diera para $t = t_0$, entonces $\partial T/\partial t$ sería no positiva. Asimismo, en este punto las derivadas segundas respecto a las variables x , y y z serían no negativas. En consecuencia, $\Delta T - (1/a^2)(\partial T/\partial t)$ sería no negativo, lo que en nuestro caso es imposible.

Exactamente de la misma forma podemos establecer que, si $\Delta T - (1/a^2)(\partial T/\partial t) > 0$, entonces en el interior de Ω , para $0 < t \leq t_0$, no puede existir un máximo para la función T .

Finalmente, si $\Delta T - (1/a^2)(\partial T/\partial t) = 0$, entonces en el interior de Ω , para $0 < t \leq t_0$, la función T no puede alcanzar un máximo ni un mínimo absolutos, ya que si T alcanzase, por ejemplo, un mínimo, entonces sumándole el término $\eta(t - t_0)$ y considerando la función $T_1 = T + \eta(t - t_0)$, no destruiríamos el mínimo absoluto con tal de que η fuera suficientemente pequeño, y entonces $\Delta T_1 - (1/a^2)(\partial T_1/\partial t)$ sería negativo, lo cual es imposible.

De la misma forma, podemos también probar la ausencia de un máximo absoluto de T en el dominio considerado.

Sin embargo, un mínimo, o un máximo, absoluto de temperatura sí puede alcanzarse en el instante inicial $t = 0$ o sobre la frontera S del medio. Si $T = 0$ en el instante inicial y sobre la frontera, entonces tenemos la identidad $T = 0$ en todo el interior del dominio para todo $t \leq t_0$. Si dos distribuciones de temperatura T_1 y T_2 tienen idénticos valores para $t = 0$ y sobre el contorno, entonces su diferencia $T_1 - T_2 = T$ satisfará la ecuación del calor y se anulará para $t = 0$ y sobre el contorno. Esto implica que $T_1 - T_2$ será igual a cero en todo punto, por lo que las dos distribuciones de temperatura T_1 y T_2 serán idénticas.

En el estudio que haremos más adelante de los métodos para resolver las ecuaciones de la física matemática veremos que el valor de T para $t = 0$ y el segundo miembro de una de las ecuaciones (12) se pueden dar arbitrariamente, es decir, que la solución de tal problema existe.

La energía de las oscilaciones y el problema de contorno para la ecuación de oscilación

Consideremos las condiciones bajo las que la tercera de las ecuaciones diferenciales básicas, la ecuación (9), tiene una única solución.

Por sencillez consideraremos la ecuación de la cuerda vibrante $\partial^2 u/\partial x^2 = (1/a^2)(\partial^2 u/\partial t^2)$, que es muy similar a la ecuación (9), diferenciándose únicamente de ella en el número de variables espaciales.

En el segundo miembro de esta ecuación aparece la cantidad $\partial^2 u / \partial t^2$, que expresa la aceleración de un punto arbitrario de la cuerda. El movimiento de cualquier sistema mecánico en el que las fuerzas —y en consecuencia las aceleraciones— vengan expresadas por medio de las coordenadas de los cuerpos móviles, está completamente determinado si se conocen las posiciones iniciales y las velocidades de todos los puntos del sistema. Así, pues, para la ecuación de la cuerda vibrante es natural especificar las posiciones y velocidades de todos los puntos en el instante inicial.

$$\begin{aligned} u|_{t=0} &= u_0(x) \\ \frac{\partial u}{\partial t} \Big|_{t=0} &= u_1(x). \end{aligned}$$

Pero, como ya se apuntó antes, en los extremos de la cuerda las fórmulas que expresan las leyes de la mecánica para puntos interiores dejan de aplicarse. Así, pues, en ambos extremos debemos imponer condiciones suplementarias. Si, por ejemplo, la cuerda está fijada en una posición de equilibrio en ambos extremos, entonces tendremos

$$u|_{x=0} = u|_{x=l} = 0.$$

Estas condiciones pueden reemplazarse a veces por otras más generales, pero un cambio de esta clase no es de importancia transcendental.

El problema de hallar las soluciones necesarias de la ecuación (9) es análogo. Para que tal solución esté bien definida, es habitual imponer las condiciones

$$\begin{aligned} p|_{t=0} &= \phi_0(x, y, z), \\ \frac{\partial p}{\partial t} \Big|_{t=0} &= \phi_1(x, y, z), \end{aligned} \quad (13)$$

y también una de las «condiciones de contorno»

$$p|_S = \phi(Q), \quad (14)$$

$$\frac{\partial p}{\partial n} \Big|_S = \psi(Q), \quad (14')$$

$$\alpha \frac{\partial p}{\partial n} \Big|_S + \beta p|_S = \chi(Q)^3. \quad (14'')$$

³ Si los segundos miembros de las condiciones (13) y (14) son nulos, tales condiciones se llaman «homogéneas».

La diferencia con respecto al caso anterior es simplemente que en vez de una sola condición inicial, la ecuación (11), tenemos ahora las dos condiciones (13).

La ecuaciones (14) expresan evidentemente las leyes físicas para las partículas sobre el contorno del volumen en cuestión.

La demostración de que en el caso general las condiciones (13), junto con una cualquiera de las condiciones (14), determinan unívocamente una solución del problema será omitida. Probaremos solamente que la solución puede ser única para una de las condiciones (14).

Supongamos que una función u verifica la ecuación

$$\frac{\Delta u}{\partial x^2} = \frac{1}{a^2} \frac{\partial^2 u}{\partial t^2},$$

con condiciones iniciales

$$u|_{t=0} = 0, \quad \frac{\partial u}{\partial t} \Big|_{t=0} = 0$$

y condición de contorno

$$\frac{\partial u}{\partial n} \Big|_S = 0.$$

(Sería igual de fácil discutir el caso en que $u|_S = 0$.)

Demostraremos que bajo estas condiciones la función u debe ser idénticamente nula.

Para probar esta propiedad no bastan los argumentos introducidos con anterioridad para establecer la unicidad de la solución del primero de ambos problemas. En cambio aquí podemos hacer uso de la interpretación física.

Sólo precisamos de una ley física, la ley de «conservación de la energía». Nos restringiremos de nuevo, por razones de simplicidad, al caso de la cuerda vibrante; el desplazamiento $u(x, t)$ de sus puntos verifica la ecuación

$$T \frac{\partial^2 u}{\partial x^2} = \rho \frac{\partial^2 u}{\partial t^2}.$$

La energía cinética de cada partícula de la cuerda, al oscilar de μ a $\mu + d\mu$, viene expresada por:

$$\frac{1}{2} \left(\frac{\partial u}{\partial t} \right)^2 \rho dx.$$

Además de energía cinética, la cuerda en su posición desplazada posee también energía potencial, creada por el incremento de longitud en comparación con la posición rectilínea. Calculemos esta energía potencial. Concentrémonos en un elemento de cuerda comprendido entre los puntos x y $x + dx$. Este elemento tiene una posición inclinada respecto al eje Ox , tal que su longitud es aproximadamente igual a

$$\sqrt{(dx)^2 + \left(\frac{\partial u}{\partial x} dx\right)^2};$$

por tanto, su alargamiento es

$$\sqrt{1 + \left(\frac{\partial u}{\partial x}\right)^2} dx - dx \approx \frac{1}{2} \left(\frac{\partial u}{\partial x}\right)^2 dx.$$

Multiplicando este alargamiento por la tensión T , hallamos la energía potencial del elemento alargado de la cuerda

$$\frac{1}{2} T \left(\frac{\partial u}{\partial x}\right)^2 dx.$$

La energía total de la cuerda de longitud l se obtiene sumando la energía cinética y potencial en todos los puntos de la cuerda. Obtenemos

$$E = \frac{1}{2} \int_0^l \left[T \left(\frac{\partial u}{\partial x}\right)^2 + \rho \left(\frac{\partial u}{\partial t}\right)^2 \right] dx.$$

Si las fuerzas que actúan sobre los extremos no efectúan trabajo, en particular si los extremos de la cuerda están fijos, entonces la energía total debe ser constante:

$$E = \text{const.}$$

Nuestra expresión para la ley de conservación de la energía es un corolario matemático de las ecuaciones básicas de la mecánica y se puede derivar de ellas. Como ya hemos escrito las leyes de movimiento en la forma de la ecuación diferencial de la cuerda vibrante con condiciones sobre los extremos, podemos dar la siguiente demostración matemática de la ley de conservación de la energía en este

caso. Si derivamos E con respecto al tiempo, y haciendo uso de reglas generales básicas, tenemos:

$$\frac{dE}{dt} = \int_0^l \left(T \frac{\partial u}{\partial x} \frac{\partial^2 u}{\partial x \partial t} + \rho \frac{\partial u}{\partial t} \frac{\partial^2 u}{\partial t^2} \right) dx.$$

Utilizando la ecuación de ondas (6) y sustituyendo $\rho(\partial^2 u / \partial t^2)$ por $T(\partial^2 u / \partial x^2)$, obtenemos dE/dt en la forma

$$\begin{aligned} \frac{dE}{dt} &= \int_0^l T \left(\frac{\partial u}{\partial x} \frac{\partial^2 u}{\partial x \partial t} + \frac{\partial u}{\partial t} \frac{\partial^2 u}{\partial x^2} \right) dx = \\ &= \int_0^l T \frac{\partial}{\partial x} \left(\frac{\partial u}{\partial x} \frac{\partial u}{\partial t} \right) dx = T \frac{\partial u}{\partial x} \frac{\partial u}{\partial t} \Big|_{x=l} - T \frac{\partial u}{\partial x} \frac{\partial u}{\partial t} \Big|_{x=0}. \end{aligned}$$

Si $(\partial u / \partial x) |_{x=0}$ o $u |_{x=0}$ es nulo, y también $(\partial u / \partial x) |_{x=l}$ o $u |_{x=l}$ es nulo, entonces

$$\frac{dE}{dt} = 0,$$

lo cual prueba que E es constante.

La ecuación de ondas (9) puede ser tratada exactamente igual para probar que la ley de conservación de la energía también es válida en este caso. Si p verifica la ecuación (9) y la condición

$$p|_S = 0 \quad \text{ó} \quad \frac{\partial p}{\partial n} \Big|_S = 0,$$

entonces la cantidad

$$E = \iiint \left[\left(\frac{\partial p}{\partial x} \right)^2 + \left(\frac{\partial p}{\partial y} \right)^2 + \left(\frac{\partial p}{\partial z} \right)^2 + \frac{1}{a^2} \left(\frac{\partial p}{\partial t} \right)^2 \right] dx dy dz$$

no dependerá de t .

Si en el instante inicial la energía total de la oscilación es igual a cero, entonces permanecerá siempre igual a cero, y esto es posible solamente en el caso de que no haya movimiento. Si el problema de integrar la ecuación de ondas con condiciones iniciales y de contorno tuviera dos soluciones p_1 y p_2 , entonces $v = p_1 - p_2$ sería una solución de la ecuación de ondas que verificaría las condiciones con cero en el segundo miembro, es decir, las condiciones homogéneas.

En este caso, cuando calculásemos la «energía» de una tal oscilación, descrita por una función v , descubriríamos que la energía $E(v)$ es igual a cero en el instante inicial. Esto significa que es siempre igual a cero y por lo tanto que la función v es idénticamente nula, con lo que las dos soluciones p_1 y p_2 son idénticas. En consecuencia, la solución del problema es única.

De esta manera llegamos al convencimiento de que los tres problemas están correctamente planteados.

Incidentalmente hemos descubierto algunas propiedades muy simples de las soluciones de estas ecuaciones. Por ejemplo, las soluciones de la ecuación de Laplace tienen la siguiente propiedad de máximo: las funciones que verifican esta ecuación alcanzan su mayor y menor valor sobre el contorno de su dominio de definición.

Las funciones que describen la distribución de calor en un medio tienen una propiedad de máximo algo diferente. Todo máximo o mínimo de temperatura en un punto se desvanece y disminuye gradualmente con el tiempo. La temperatura en un punto puede aumentar o disminuir solamente si es menor o mayor que la de los puntos contiguos. La temperatura se uniformiza con el paso del tiempo. Todas las irregularidades se nivelan por el paso del calor de los lugares más calientes a los más fríos.

Por el contrario, en la propagación de las oscilaciones aquí no se produce ningún proceso de nivelación. Estas oscilaciones no disminuyen ni se nivelan, ya que la suma de sus energías cinética y potencial debe permanecer siempre constante.

§4. La propagación de ondas

Las propiedades de las oscilaciones se pueden evidenciar muy claramente con ejemplos sencillísimos. Consideremos dos casos característicos.

Nuestro primer ejemplo es la ecuación de la cuerda vibrante

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{a^2} \frac{\partial^2 u}{\partial t^2}. \quad (15)$$

Esta ecuación, como se puede comprobar, tiene dos soluciones particulares de la forma

$$u_1 = \phi_1(x - at), \quad u_2 = \phi_2(x + at),$$

donde ϕ_1 y ϕ_2 son funciones arbitrarias dos veces diferenciables.

Por derivación directa es fácil probar que las funciones u_1 y u_2 verifican la ecuación (15). Se puede demostrar que

$$u = u_1 \div u_2$$

es una solución general de esta ecuación.

La forma general de las oscilaciones descritas por las funciones u_1 y u_2 es de considerable interés. Para considerarla de la manera más conveniente, hagamos mentalmente el siguiente experimento. Supongamos que el observador de la cuerda vibrante no es estacionario sino que se mueve a lo largo del eje Ox con velocidad a . Para un tal observador la posición de un punto de la cuerda estará determinada, no por un sistema de coordenadas estacionario, sino móvil. Sea ξ la coordenada x de este sistema. Entonces $\xi = 0$ corresponderá evidentemente en cada instante al valor $x = at$. Por tanto está claro que

$$\xi = x - at.$$

Podemos representar una función arbitraria $u(x, t)$ en la forma

$$u(x, t) = \phi(\xi, t).$$

Para la solución u_1 tendremos

$$u_1(x, t) = \phi_1(\xi),$$

con lo que en este sistema de coordenadas la solución $u_1(x, t)$ resulta ser independiente del tiempo. En consecuencia, un observador que se mueva con velocidad a ve la cuerda como una curva estacionaria. Un observador estacionario, por el contrario, ve en la cuerda una onda que se mueve a lo largo del eje Ox con velocidad a .

Análogamente, la solución $u_2(x, t)$ puede ser considerada como una onda que discurre en dirección opuesta con velocidad a . Si la cuerda fuese infinita, ambas ondas se propagarían infinitamente lejos. Al moverse en direcciones opuestas, pueden producir, por superposición, formas muy curiosas en la cuerda. El desplazamiento resultante puede aumentar en unas ocasiones y disminuir en otras.

Si, al llegar a un punto dado desde lados opuestos, u_1 y u_2 tienen el mismo signo, entonces se refuerzan mutuamente; pero, si tienen signos opuestos, se contrarrestan. La figura 3 muestra varias posiciones sucesivas de la cuerda para dos desplazamientos particulares.

Inicialmente las ondas se mueven independientemente una hacia la otra, para después comenzar a interactuar. En el segundo caso de la figura 3 habrá un instante de completo aniquilamiento de las oscilaciones, tras lo cual las ondas se volverán a separar.

Otro ejemplo que se presta fácilmente a un estudio cualitativo es la propagación de ondas en el espacio.

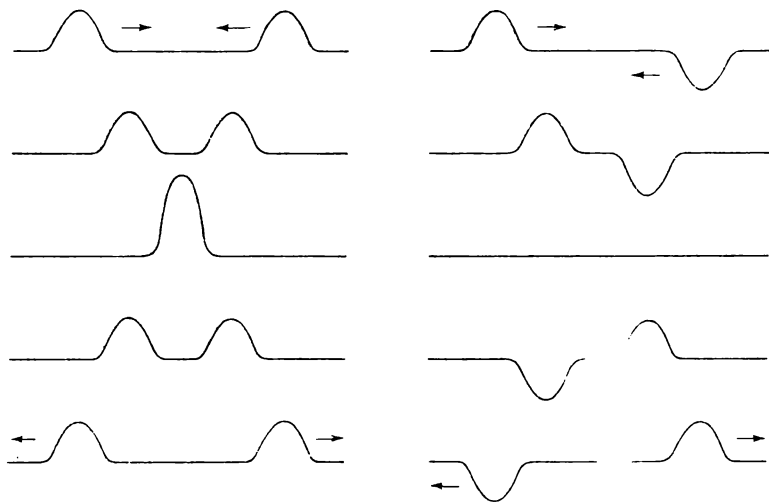


FIG. 3

La ecuación

$$\Delta u = \frac{1}{a^2} \frac{\partial^2 u}{\partial t^2}, \quad (16)$$

obtenida antes, tiene dos soluciones particulares de la forma

$$u_1 = \frac{1}{r} \phi_1(r - at), \quad u_2 = \frac{1}{r} \phi_2(r + at), \quad (17)$$

donde r denota la distancia de un punto dado al origen del sistema de coordenadas, $r^2 = x^2 + y^2 + z^2$, y ϕ_1 y ϕ_2 son funciones arbitrarias, dos veces diferenciables.

La demostración de que u_1 y u_2 son soluciones sería demasiado prolija y la omitimos aquí.

La forma de las ondas descritas por estas soluciones es en general la misma que para la cuerda. Si prescindimos del factor $1/r$ que aparece en el segundo miembro, entonces la primera solución representa una onda que se propaga en la dirección de las r crecientes. Esta onda tiene simetría esférica; es idéntica en todos los puntos que tienen el mismo valor de r .

El factor $1/r$ determina el que la amplitud de la onda sea inversamente proporcional a la distancia desde el origen. Tal oscilación se denomina onda esférica divergente. Una buena imagen de ello son los círculos que se extienden sobre la superficie del agua cuando se arroja una piedra en ella, sólo que en este caso las ondas son circulares en vez de esféricas.

Esta segunda solución de (17) es también de gran interés; se denomina onda convergente y se propaga en dirección al origen. Su amplitud tiende a infinito a medida que se aproxima al origen. Vemos que una concentración tal de la perturbación en un punto puede conducir, aun cuando las oscilaciones iniciales sean pequeñas, a un choque brutal.

§5. Métodos de construcción de soluciones

Sobre la posibilidad de descomponer una solución en otras más simples

Las soluciones de los problemas de la física matemática formulados con anterioridad pueden ser obtenidas mediante diversos artificios, que constituyen problemas específicos distintos. No obstante, en la base de estos métodos hay una idea general. Como hemos visto, todas las ecuaciones de la física matemática son, para pequeños valores de las funciones incógnitas, lineales con respecto a éstas y a sus derivadas. Las condiciones de contorno y las iniciales son también lineales.

Si formamos la diferencia entre dos soluciones cualesquiera de la misma ecuación, dicha diferencia será también solución de la ecuación con el segundo miembro igual a cero. Tal ecuación se denomina ecuación homogénea. Por ejemplo, para la ecuación de Poisson $\Delta u = -4\pi\rho$, la homogénea correspondiente es la ecuación de Laplace $\Delta u = 0$.

Si dos soluciones de la misma ecuación verifican también las mismas condiciones de contorno, su diferencia satisfará la correspon-

diente condición homogénea: los valores de la expresión correspondiente sobre el contorno serán iguales a cero.

Por tanto, la variedad total de las soluciones de tal ecuación, para condiciones de contorno dadas, se puede hallar tomando, por una parte, una solución particular que verifique las condiciones no homogéneas dadas, y por otra, todas las posibles soluciones de la ecuación homogénea que verifique las condiciones de contorno homogéneas (pero no, en general, las iniciales).

Las soluciones de ecuaciones homogéneas que verifican condiciones de contorno homogéneas se pueden sumar o multiplicar por constantes sin que dejen de ser soluciones.

Si una solución de una ecuación homogénea con condiciones homogéneas es función de un parámetro, integrando respecto a éste obtendremos también una tal solución. Estos hechos forman la base del método más importante de resolución de problemas lineales de todas clases para las ecuaciones de la física matemática: el método de superposición.

La solución del problema será de la forma

$$u = u_0 + \sum_k u_k,$$

donde u_0 es una solución particular de la ecuación, solución que verifica las condiciones de contorno, aunque no las iniciales; las u_k son soluciones de la correspondiente ecuación homogénea, y verifican las correspondientes condiciones de contorno homogéneas. Si la ecuación y las condiciones de contorno fueran originalmente homogéneas, entonces la solución del problema se habrá de buscar en la forma

$$u = \sum u_k.$$

Con el fin de poder satisfacer condiciones iniciales arbitrarias mediante la elección de soluciones particulares u_k de la ecuación homogénea, es preciso disponer de un arsenal suficientemente grande de tales soluciones.

El método de separación de variables

Para la construcción de este arsenal de soluciones existe un método denominado de *separación de variables* o *método de Fourier*.

Examinemos este método, por ejemplo, para resolver el problema

$$\Delta u = \frac{\partial^2 u}{\partial t^2}, \quad (18)$$

$$u|_S = 0, \quad u|_{t=0} = f_0(x, y, z), \quad u'_t|_{t=0} = f_1(x, y, z).$$

Para buscar una solución particular de la ecuación, supongamos en primer lugar que la función deseada u verifica la condición de contorno $u|_S = 0$ y que puede expresarse como producto de dos funciones, una de las cuales depende solamente del tiempo t , y la otra, sólo de las variables espaciales:

$$u(x, y, z, t) = U(x, y, z) T(t).$$

Sustituyendo esta presunta solución en nuestra ecuación, tenemos

$$T(t) \Delta U = T''(t) U.$$

Dividiendo ambos miembros por TU se obtiene

$$\frac{T''}{T} = \frac{\Delta U}{U}.$$

El segundo miembro de esta ecuación es sólo función de las variables espaciales, mientras que el primero es independiente de las mismas. Por tanto, se deduce que la ecuación dada sólo puede verificarse si ambos miembros son iguales al mismo valor constante. Esto nos conduce a un sistema de dos ecuaciones

$$\frac{T''}{T} = -\lambda_k^2, \quad \frac{\Delta U}{U} = -\lambda_k^2.$$

La cantidad constante de la derecha se denota por $-\lambda_k^2$ para poner de relieve que es negativa (lo cual se puede probar rigurosamente). El subíndice k se introduce para hacer notar que existen infinitos valores posibles de $-\lambda_k^2$; las soluciones correspondientes a éstos forman un sistema de funciones que es completo en un sentido bien conocido.

Por multiplicación cruzada en ambas ecuaciones, obtenemos

$$T'' + \lambda_k^2 T = 0; \quad \Delta U + \lambda_k^2 U = 0.$$

La primera de estas ecuaciones tiene, como sabemos, la solución

$$T = A_k \cos \lambda_k t + B_k \operatorname{sen} \lambda_k t,$$

donde A_k y B_k son constantes arbitrarias. Esta solución se puede simplificar introduciendo el argumento auxiliar ϕ . Tenemos

$$\frac{A_k}{\sqrt{A_k^2 + B_k^2}} = \operatorname{sen} \phi_k, \quad \frac{B_k}{\sqrt{A_k^2 + B_k^2}} = \cos \phi_k, \quad \sqrt{A_k^2 + B_k^2} = M_k.$$

Entonces

$$T = \sqrt{A_k^2 + B_k^2} \operatorname{sen} (\lambda_k t + \phi_k) = M_k \operatorname{sen} (\lambda_k t + \phi_k).$$

La función T representa una oscilación armónica de frecuencia λ_k y ángulo de fase ϕ_k .

Más difícil e interesante es el problema de hallar una solución de la ecuación

$$\Delta U + \lambda_k^2 U = 0 \tag{19}$$

para condiciones de contorno homogéneas dadas; por ejemplo, para la condición

$$U|_S = 0$$

(donde S es la frontera del volumen Ω en cuestión), o para cualquier otra condición homogénea. La solución de este problema no es siempre fácil de construir como combinación finita de funciones conocidas, aunque siempre existe y puede hallarse con el grado deseado de aproximación.

La ecuación $\Delta U + \lambda_k^2 U = 0$ con la condición $U|_S = 0$ tiene en principio la solución evidente $U \equiv 0$. Esta solución es trivial y completamente inútil para nuestro propósito. Si los λ_k son números elegidos arbitrariamente, entonces en general el problema no tendrá solución. Sin embargo, existen normalmente valores de λ_k para los cuales la ecuación sí tiene una solución no trivial.

Todos los valores posibles de la constante λ_k^2 están determinados por el requisito de que la ecuación (19) tenga una solución no trivial (es decir, distinta de la función idénticamente nula) que satisfaga la condición $U|_S = 0$. De esto se sigue también que los números denotados por $-\lambda_k^2$ deben ser negativos.

Para cada uno de los valores posibles de λ_k en la ecuación (19)

podemos hallar al menos una función U_k . Esto nos permite construir una solución particular de la ecuación de ondas (18) en la forma

$$u_k = M_k \operatorname{sen}(\lambda_k t + \phi_k) U_k(x, y, z).$$

Tal solución se denomina *oscilación característica* (u *oscilación propia*) del volumen considerado. La constante λ_k es la frecuencia de la oscilación característica, y la función $U_k(x, y, z)$ determina su forma. Esta función se denomina por lo común *función propia* (o *función característica*, o *autofunción*). En todo instante, la función u_k , considerada como función de las variables x , y y z , diferirá de la función $U_k(x, y, z)$ solamente en la escala.

No disponemos de espacio aquí para una demostración detallada de las muchas propiedades notables de las oscilaciones y funciones características; por consiguiente nos restringiremos solamente a enumerar algunas de ellas.

La primera propiedad de las oscilaciones características consiste en el hecho de que para un volumen dado existe un conjunto numerable de frecuencias características. Estas frecuencias tienden a infinito cuando crece k .

Otra propiedad de las oscilaciones características se denomina *ortogonalidad*. Consiste en el hecho de que la integral sobre el dominio Ω del producto de las funciones características correspondientes a distintos valores de λ_k es igual a cero ⁴.

$$\iiint_{\Omega} U_k(x, y, z) U_j(x, y, z) dx dy dz = 0 \quad (j \neq k).$$

Para $j = k$ supondremos

$$\iiint_{\Omega} U_k(x, y, z)^2 dx dy dz = 1.$$

Esto siempre se puede conseguir multiplicando las funciones $U_k(x, y, z)$ por una constante apropiada, cuya elección no altera el hecho de que la función satisfaga la ecuación (19) y la condición $U|_S = 0$.

⁴ Si a un mismo valor de λ corresponden varias funciones U esencialmente diferentes (linealmente independientes), entonces este valor de λ se considera como si apareciese un número correspondiente de veces en el conjunto de valores característicos λ_k . La condición de ortogonalidad para funciones correspondientes al mismo valor de λ_k se puede garantizar mediante una elección adecuada de estas funciones.

Finalmente, una tercera propiedad de las oscilaciones características es que, si no excluimos ningún valor de λ_k , entonces mediante las funciones características $U_k(x, y, z)$ podemos representar, con cualquier grado de aproximación deseado, una función completamente arbitraria $f(x, y, z)$, con tal solamente de que ésta verifique la condición de contorno $f|_S = 0$ y tenga la primera y segunda derivadas continuas. Tal función $f(x, y, z)$ se puede representar mediante la serie convergente

$$f(x, y, z) = \sum_{k=1}^{\infty} C_k U_k(x, y, z). \quad (20)$$

La tercera propiedad de las funciones características nos proporciona en principio la posibilidad de representar una función $f(x, y, z)$ mediante una serie de funciones características de nuestro problema, y aplicando la segunda propiedad podemos calcular todos los coeficientes de esta serie. En efecto, si multiplicamos ambos miembros de la ecuación (20) por $U_j(x, y, z)$ e integramos sobre el dominio Ω , obtenemos

$$\begin{aligned} \iiint_{\Omega} f(x, y, z) U_j(x, y, z) dx dy dz &= \\ &= \sum_{k=1}^{\infty} C_k \iiint_{\Omega} U_k(x, y, z) U_j(x, y, z) dx dy dz. \end{aligned}$$

En la suma del segundo miembro, todos los términos para los que $k \neq j$ desaparecen en virtud de la ortogonalidad, y el coeficiente de C_j es igual a uno. En consecuencia tenemos

$$C_j = \iiint_{\Omega} f(x, y, z) U_j(x, y, z) dx dy dz.$$

Estas propiedades de las oscilaciones características nos permiten ahora resolver el problema general de las oscilaciones para unas condiciones iniciales cualesquiera.

Para ello supongamos que tenemos una solución del problema en la forma

$$u = \sum U_k(x, y, z) (A_k \cos \lambda_k t + B_k \sin \lambda_k t) \quad (21)$$

y que tratamos de elegir las constantes A_k y B_k de tal manera que tengamos

$$u|_{t=0} = f_0(x, y, z),$$

$$\left. \frac{\partial u}{\partial t} \right|_{t=0} = f_1(x, y, z).$$

Poniendo $t = 0$ en el segundo miembro de (21), vemos que los términos en seno desaparecen y $\cos \lambda_k t$ se hace igual a uno, así que tendremos

$$f_0(x, y, z) = \sum_{k=1}^{\infty} A_k U_k(x, y, z).$$

Por la tercera propiedad, las oscilaciones características se pueden utilizar para tal representación; y por la segunda propiedad tenemos

$$A_k = \iiint_{\Omega} f_0(x, y, z) U_k(x, y, z) dx dy dz.$$

Análogamente, derivando la fórmula (21) con respecto a t y poniendo $t = 0$, tendremos

$$\begin{aligned} \left. \frac{\partial u}{\partial t} \right|_{t=0} &= f_1(x, y, z) = \sum_{k=1}^{\infty} \lambda_k (B_k \cos \lambda_k t - A_k \sin \lambda_k t)|_{t=0} U_k(x, y, z) = \\ &= \sum_{k=1}^{\infty} \lambda_k B_k U_k(x, y, z). \end{aligned}$$

Así, pues, como antes, obtenemos los valores de B_k de la forma siguiente:

$$B_k = \frac{1}{\lambda_k} \iiint_{\Omega} f_1(x, y, z) U_k(x, y, z) dx dy dz.$$

Conocidos A_k y B_k , conocemos las fases y las amplitudes de todas las oscilaciones características.

De esta manera hemos probado que por adición de oscilaciones características es posible obtener la solución más general del problema con condiciones de contorno homogéneas.

Toda solución, pues, consta de oscilaciones características, cuya amplitud y fase se pueden calcular conociendo las condiciones iniciales.

Exactamente de la misma manera podemos estudiar oscilaciones con un menor número de variables independientes. Como ejemplo consideremos la cuerda vibrante, fija por los extremos. La ecuación de la cuerda vibrante tiene la forma

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}.$$

Supongamos que queremos calcular una solución del problema para una cuerda de longitud l fija por los extremos:

$$u|_{x=0} = u|_{x=l} = 0.$$

Buscamos una colección de soluciones particulares

$$u_k = T_k(t) U_k(x).$$

Evidentemente obtenemos, como antes,

$$T_k'' U_k = a^2 U_k'' T_k,$$

o bien

$$\frac{T_k''}{T_k} = a^2 \frac{U_k''}{U_k} = -\lambda_k^2.$$

De donde

$$T_k = A_k \cos \lambda_k t + B_k \sin \lambda_k t,$$

$$U_k = M_k \cos \frac{\lambda_k}{a} x + N_k \sin \frac{\lambda_k}{a} x.$$

Utilicemos las condiciones de contorno para calcular los valores de λ_k . Para λ_k arbitrario no es posible verificar ambas condiciones de contorno. De la condición $U_k|_{x=0} = 0$ obtenemos $M_k = 0$, y esto significa que $U_k = N_k \sin (\lambda_k/a)x$. Poniendo $x = l$, obtenemos $\sin (\lambda_k l/a) = 0$. Esto solamente puede suceder si $\lambda_k l/a = k\pi$, donde k es un entero. Esto significa que

$$\lambda_k = \frac{ak\pi}{l}.$$

La condición $\int_0^l U_k^2 dx = 1$ prueba que $N_k = \sqrt{\frac{2}{l}}$. Finalmente

$$U_k(x) = \sqrt{\frac{2}{l}} \sin \frac{k\pi x}{l}, \quad T_k = A_k \cos \frac{ak\pi t}{l} + B_k \sin \frac{ak\pi t}{l}.$$

De esta manera las oscilaciones características de la cuerda, como vemos, tienen forma sinusoidal con un número entero de semiondas sobre el total de la cuerda. Cada oscilación tiene su propia frecuencia, y éstas se pueden clasificar en orden creciente

$$\frac{a\pi}{l}, 2 \frac{a\pi}{l}, 3 \frac{a\pi}{l}, \dots, k \frac{a\pi}{l}, \dots.$$

Es bien conocido que estas frecuencias son precisamente las que oímos en las vibraciones de una cuerda sonora. La frecuencia $a\pi/l$ se denomina *frecuencia fundamental*, y las frecuencias restantes son *armónicos*. Las funciones características $\sqrt{2/l} \sin(k\pi x/l)$ sobre el intervalo $0 \leq x \leq l$ cambian de signo $k-1$ veces, ya que $k\pi x/l$ recorre todos los valores de 0 a $k\pi$, lo cual significa que su seno cambia de signo $k-1$ veces. Los puntos en que las funciones características U_k se anulan se llaman *nodos* de las oscilaciones.

Si conseguimos de alguna manera que la cuerda no se mueva en un punto correspondiente a un nodo, por ejemplo, del primer armónico, entonces el tono fundamental desaparecerá y oiremos solamente el sonido del primer armónico, que es una octava más alto. Tal artificio, llamado *parada*, se utiliza en instrumentos que se tocan con un arco: el violín, la viola y el violonchelo.

Hemos analizado el método de separación de variables en tanto que aplicado al problema de hallar oscilaciones características. No obstante, el método se puede aplicar con mucha mayor amplitud a los problemas de flujo de calor y a toda una serie de otros problemas.

Para la ecuación del flujo de calor

$$\Delta T = \frac{\partial T}{\partial t}$$

con la condición

$$T|_S = 0$$

tendremos, como antes,

$$T = \sum F_k(t) U_k(x, y, z).$$

Aquí

$$\frac{F'_k(t)}{F_k(t)} = -\lambda_k^2, \quad \Delta U_k + \lambda_k^2 U_k = 0.$$

La solución se obtiene en la forma

$$T = \sum_{k=1}^{\infty} e^{-\lambda_k^2 t} U_k(x, y, z).$$

Este método ha sido también usado con gran éxito para resolver algunas otras ecuaciones. Consideremos, por ejemplo, la ecuación de Laplace

$$\Delta u = 0$$

en el círculo

$$x^2 + y^2 \leq 1,$$

y supongamos que hemos de construir una solución que satisfaga la condición

$$u|_{r=1} = f(\vartheta),$$

donde r y ϑ denotan las coordenadas polares de un punto en el plano.

La ecuación de Laplace se puede transformar fácilmente en coordenadas polares, adoptando la forma

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \vartheta^2} = 0.$$

Queremos hallar una solución de esta ecuación en la forma

$$u = \sum_{k=1}^{\infty} R_k(r) \theta_k(\vartheta).$$

Si imponemos que cada término de la serie verifique individualmente la ecuación, tenemos

$$\left[R_k''(r) + \frac{1}{r} R_k'(r) \right] \theta_k(\vartheta) + \frac{1}{r^2} \theta_k''(\vartheta) R_k(r) = 0.$$

Dividiendo la ecuación por $R_k(r) \theta_k(\vartheta)/r^2$, obtenemos

$$\frac{r^2 \left[R_k''(r) + \frac{1}{r} R_k'(r) \right]}{R_k(r)} = - \frac{\theta_k''(\vartheta)}{\theta_k(\vartheta)}$$

Poniendo nuevamente

$$\frac{\theta_k''(\vartheta)}{\theta_k(\vartheta)} = -\lambda_k^2,$$

tenemos

$$r^2 \left[R_k'' + \frac{1}{r} R_k' \right] - \lambda_k^2 R_k = 0.$$

Es fácil ver que la función $\theta_k(\vartheta)$ debe ser una función periódica de ϑ con período 2π . Integrando la ecuación $\theta_k''(\vartheta) + \lambda_k^2 \theta_k(\vartheta) = 0$, obtenemos

$$\theta_k = a_k \cos \lambda_k \vartheta + b_k \operatorname{sen} \lambda_k \vartheta.$$

Esta función será periódica con el período requerido solamente si λ_k es entero. Poniendo $\lambda_k = k$, tenemos

$$\theta_k = a_k \cos k \vartheta + b_k \operatorname{sen} k \vartheta.$$

La ecuación para R_k tiene una solución general de la forma

$$R_k = A r^k + \frac{B}{r^k}.$$

Reteniendo solamente el término que está acotado para $r \rightarrow 0$, obtenemos la solución general de la ecuación de Laplace en la forma

$$u = a_0 + \sum_{k=1}^{\infty} (a_k \cos k \vartheta + b_k \operatorname{sen} k \vartheta) r^k.$$

Este método puede usarse normalmente para calcular soluciones no triviales de la ecuación $\Delta U_k + \lambda_k^2 U_k = 0$ que verifiquen condiciones de contorno homogéneas. Cuando el problema pueda reducirse al de resolver ecuaciones diferenciales ordinarias, decimos que per-

mite una separación completa de variables. Esta separación completa de variables puede hacerse, como fue probado por el matemático soviético V. V. Stepanov, solamente en algunos casos especiales. El método de separación de variables era conocido por los matemáticos hace mucho tiempo. Fue utilizado esencialmente por Euler, Bernoulli y d'Alembert. Fourier lo usó sistemáticamente para la solución de problemas de física matemática, particularmente en la conducción del calor. Sin embargo, como ya hemos dicho, este método es a menudo inaplicable; debemos utilizar otros procedimientos, que pasamos a discutir ahora.

El método de los potenciales

La idea esencial de este método es, como antes, la superposición de soluciones particulares para la construcción de una solución en forma general. Pero esta vez para las soluciones particulares fundamentales usaremos funciones que se hagan infinitas en un punto. Ilustrémoslo con las ecuaciones de Laplace y Poisson.

Sea M_0 un punto de nuestro espacio. Denotemos por $r(M, M_0)$ la distancia del punto M_0 al punto variable M . La función $1/r(M, M_0)$ para M_0 fijo es una función del punto variable M . Es fácil establecer el hecho de que esta función es una función armónica del punto M en todo el espacio ⁵, excepto, por supuesto, en el punto M_0 , donde se hace infinita, junto con sus derivadas.

La suma de varias funciones de esta forma

$$\sum_{i=1}^N A_i \frac{1}{r(M, M_i)},$$

donde los puntos $M_1, M_2, \dots, M_N$ son puntos cualesquiera del espacio, es también una función armónica del punto M . Esta función tendrá singularidades en todos los puntos M_i . Si elegimos los puntos $M_1, M_2, \dots, M_N$ tan densamente distribuidos como queramos en un volumen Ω , y al mismo tiempo multiplicamos por coeficientes A_i , podemos pasar al límite en esta expresión y obtener una nueva función

$$U = \lim \sum_{i=1}^N \frac{A_i}{r(M, M_i)} = \int \int_{\Omega} \frac{A(M')}{r(M, M')} d\Omega,$$

⁵ Esto es, la función verifica la ecuación de Laplace.

donde el punto M' recorre todo el volumen Ω . La integral en esta forma se denomina *potencial newtoniano*. Se puede probar, aunque no lo haremos aquí, que la función U así construida verifica la ecuación $\Delta U = -4\pi A$.

El potencial newtoniano tiene un significado físico sencillo. Para comprenderlo, comenzaremos con la función $A_i/r(M, M_i)$.

Las derivadas parciales de esta función con respecto a las coordenadas son

$$A_i \frac{x_i - x}{r^3} = X, \quad A_i \frac{y_i - y}{r^3} = Y, \quad A_i \frac{z_i - z}{r^3} = Z.$$

En el punto M_i colocamos una masa A_i , la cual atraerá a todos los cuerpos con una fuerza dirigida hacia el punto M_i e inversamente proporcional al cuadrado de la distancia desde M_i . Descomponemos esta fuerza en sus componentes según los ejes de coordenadas. Si la magnitud de la fuerza que actúa sobre un punto material de masa unidad es A_i/r^2 , los cosenos de los ángulos formados por la dirección de esta fuerza y los ejes coordenados serán $(x_i - x)/r$, $(y_i - y)/r$, $(z_i - z)/r$. Por tanto, las componentes de la fuerza ejercida sobre la unidad de masa en el punto M por un centro de atracción M_i serán iguales a X , Y y Z , las derivadas parciales de la función A_i/r con respecto a las coordenadas. Si colocamos masas en los puntos $M_1, M_2, \dots, M_N$, entonces todo punto material de masa unidad situado en M será atraído por una fuerza igual a la resultante de todas las fuerzas que actúan sobre él desde los puntos dados M_i . Dicho con otras palabras:

$$X = \frac{\partial}{\partial x} \sum \frac{A_i}{r(M, M_i)}, \quad Y = \frac{\partial}{\partial y} \sum \frac{A_i}{r(M, M_i)}, \quad Z = \frac{\partial}{\partial z} \sum \frac{A_i}{r(M, M_i)}.$$

Pasando al límite y sustituyendo la suma por una integral, obtenemos

$$X = \frac{\partial U}{\partial x}, \quad Y = \frac{\partial U}{\partial y}, \quad Z = \frac{\partial U}{\partial z}, \quad \text{donde} \quad U = \iiint_{\Omega} \frac{A}{r} d\Omega.$$

La función U , con derivadas parciales iguales a las componentes de la fuerza que actúa sobre un punto, se llama *potencial* de la fuerza. Por tanto, la función $A_i/r(M, M_i)$ es el potencial de la atracción ejercida por el punto M_i ; la función $\Sigma [A_i/r(M, M_i)]$ es el potencial de la atracción ejercida por el conjunto de puntos $M_1, M_2, \dots, M_N$, y la función $U = \iiint_{\Omega} (A/r) d\Omega$ es el potencial de la atracción ejercida por las masas distribuidas continuamente en el volumen Ω .

En vez de distribuir las masas en un volumen, podemos situar los puntos $M_1, M_2, \dots, M_N$ sobre una superficie S . De nuevo, al crecer el número de estos puntos, obtenemos en el límite la integral

$$V = \iint_S \frac{A(Q)}{r} ds, \quad (22)$$

donde Q es un punto de la superficie S .

No es difícil ver que esta función es armónica en el interior y en el exterior de la superficie S . Sobre la superficie la función es continua, como se puede probar, aunque sus derivadas parciales de primer orden tengan discontinuidades finitas.

Las funciones $\partial(1/r)/\partial x_i$, $\partial(1/r)/\partial y_i$ y $\partial(1/r)/\partial z_i$ también son funciones armónicas del punto M para M_i fijo. A partir de estas funciones podemos formar las sumas

$$\sum A_i \frac{\partial \frac{1}{r}}{\partial x_i} + \sum B_i \frac{\partial \frac{1}{r}}{\partial y_i} + \sum C_i \frac{\partial \frac{1}{r}}{\partial z_i},$$

que serán funciones armónicas en todos los puntos, salvo quizá en $M_1, M_2, \dots, M_N$.

De importancia particular es la integral

$$\begin{aligned} W &= \iint_S \mu(Q) \left[\frac{\partial \frac{1}{r}}{\partial x'} \cos(\mathbf{n}, \mathbf{x}) + \frac{\partial \frac{1}{r}}{\partial y'} \cos(\mathbf{n}, \mathbf{y}) + \frac{\partial \frac{1}{r}}{\partial z'} \cos(\mathbf{n}, \mathbf{z}) \right] ds = \\ &= \iint_S \mu(Q) K(Q, M) ds, \end{aligned} \quad (23)$$

en la cual x', y' y z' son las coordenadas de un punto variable Q sobre la superficie S ; $\mathbf{n}$ es la dirección de la normal a la superficie S en el punto Q ; x, y y z son las direcciones de los ejes de coordenadas, y r es la distancia de Q al punto M en el que está definido el valor de la función W .

La integral (22) se denomina *potencial de una cara simple*, y la integral (23), *potencial de una cara doble* ⁶. Los potenciales de una

⁶ Los nombres de estos potenciales están relacionados con el hecho físico siguiente. Suponemos que sobre la superficie S hemos introducido cargas eléctricas

cara doble y de una cara simple representan funciones armónicas dentro y fuera de la superficie S .

Muchos problemas en la teoría de funciones armónicas se pueden resolver utilizando potenciales. Utilizando el potencial de una cara doble, podemos resolver el problema de construir, en un dominio dado, una función armónica u que tome valores dados $2\pi\phi(Q)$ sobre la frontera S del dominio. Para construir tal función, sólo necesitamos elegir la función $\mu(Q)$ de una manera conveniente.

Este problema recuerda algo al problema similar de hallar los coeficientes de la serie

$$\phi = \sum a_k U_k$$

para que represente la función del primer miembro.

Una notable propiedad de la integral W consiste en el hecho de que su límite, cuando el punto M se aproxima a Q_0 desde el interior de la superficie, tiene la forma

$$\lim_{M \rightarrow Q_0} W = 2\pi\mu(Q_0) + \iint_S K(Q, Q_0) \mu(Q) ds.$$

Igualando esta expresión a la función dada $2\pi\phi(Q_0)$, obtenemos la ecuación

$$\mu(Q_0) + \frac{1}{2\pi} \iint_S K(Q, Q_0) \mu(Q) ds = \phi(Q_0).$$

Esta es una *ecuación integral de segunda especie*. La teoría de tales ecuaciones ha sido desarrollada por muchos matemáticos. Si logramos resolver esta ecuación por algún método, obtendremos una solución de nuestro problema original.

Análogamente podemos hallar una solución a otros problemas de la teoría de funciones armónicas. Después de elegir un potencial

cas. Estas crean en el espacio un campo eléctrico, cuyo potencial estará representado por la integral (22), que se denomina por tanto potencial de una cara simple.

Suponemos ahora que la superficie S es una delgada película no conductora. Sobre una de sus caras distribuimos, de acuerdo con cierta ley, cargas eléctricas de un signo (por ejemplo, positivo). Sobre la otra cara de S distribuimos, según la misma ley, cargas eléctricas de signo opuesto. La acción de estas dos caras eléctricas también engendra en el espacio un campo eléctrico. Como se puede calcular, el potencial de este campo estará representado por la integral (23).

conveniente, la densidad, es decir, el valor de una función arbitraria que aparezca en él, se define de tal forma que todas las condiciones impuestas se verifiquen.

Desde un punto de vista físico, esto significa que toda función armónica puede ser representada como el potencial de una doble cara eléctrica, si distribuimos esta cara sobre una superficie S con una densidad apropiada.

Construcción aproximada de soluciones; método de Galerkin y método de redes

1. Hemos discutido dos métodos para resolver ecuaciones de física matemática: el de separación completa de variables y el de potenciales. Estos métodos fueron desarrollados por los científicos de los siglos XVIII y XIX, Fourier, Poisson, Ostrogradski, Liapunov y otros. En el siglo XX se han añadido otros muchos métodos. Examinaremos dos de ellos, el de Galerkin y el de diferencias finitas, o método de redes.

El primero fue propuesto por el académico B. G. Galerkin para resolver ecuaciones de la forma

$$\sum \sum \sum \sum A_{ijkl} \frac{\partial^4 U}{\partial x_i \partial x_j \partial x_k \partial x_l} + \sum \sum \sum B_{ijk} \frac{\partial^3 U}{\partial x_i \partial x_j \partial x_k} + \\ + \sum \sum C_{ij} \frac{\partial^2 U}{\partial x_i \partial x_j} + \sum D_i \frac{\partial U}{\partial x_i} + EU + \lambda U = 0,$$

que contienen un parámetro desconocido λ , donde los índices i, j, k y l toman independientemente los valores 1, 2 y 3. Estas ecuaciones derivan de las ecuaciones que contienen una variable independiente t utilizando el método de separación de variables, de la misma forma que la ecuación de ondas

$$\Delta u = \frac{\partial^2 u}{\partial t^2}$$

conduce a la ecuación $\Delta U + \lambda^2 U = 0$. El problema consiste en hallar aquellos valores de λ para los cuales el problema de contorno homogéneo tiene una solución no nula, y construir esta solución.

La esencia del método de Galerkin es la siguiente. La función incógnita se busca en la forma aproximada

$$U \approx \sum_{m=1}^N a_m \omega_m(x_1, x_2, x_3),$$

donde $\omega_m(x_1, x_2, x_3)$ son funciones arbitrarias que verifican las condiciones de contorno.

La supuesta solución se sustituye en el primer miembro de la ecuación, resultando la ecuación aproximada

$$\sum_{m=1}^N a_m \left[\sum \sum \sum \sum A_{ijkl} \frac{\partial^4 \omega_m}{\partial x_i \partial x_j \partial x_k \partial x_l} + \sum \sum \sum B_{ijk} \frac{\partial^3 \omega_m}{\partial x_i \partial x_j \partial x_k} + \right. \\ \left. + \sum \sum C_{ij} \frac{\partial^2 \omega_m}{\partial x_i \partial x_j} + \sum D_i \frac{\partial \omega_m}{\partial x_i} + E \omega_m \right] + \lambda \sum_{m=1}^N a_m \omega_m \approx 0.$$

Por razones de brevedad denotamos la expresión dentro de los corchetes por $L\omega_m$, y escribimos la ecuación en la forma

$$\sum a_m L\omega_m + \lambda \sum a_m \omega_m \approx 0.$$

Ahora multiplicamos ambos miembros de nuestra ecuación aproximada por ω_n e integramos sobre el dominio Ω en el cual se busca la solución. Obtenemos

$$\iint_{\Omega} \sum a_m \omega_n L\omega_m d\Omega + \lambda \iint_{\Omega} \sum a_m \omega_m \omega_n d\Omega \approx 0,$$

la cual se puede volver a escribir en la forma

$$\sum_{m=1}^N a_m \iint_{\Omega} \omega_n L\omega_m d\Omega + \lambda \sum_{m=1}^N a_m \iint_{\Omega} \omega_m \omega_n d\Omega \approx 0.$$

Si nos fijamos el propósito de satisfacer estas ecuaciones exactamente, tendremos un sistema de ecuaciones algebraicas de primer grado en los coeficientes a_m . El número de ecuaciones del sistema será igual al de incógnitas, por lo cual este sistema tendrá una solución no nula sólo si su determinante es cero. Desarrollando este determinante, obtendremos una ecuación de grado N en el número desconocido λ .

Después de hallar el valor y sustituirlo en el sistema, resolvemos éste para obtener expresiones aproximadas de la función U .

El método de Galerkin es aconsejable para ecuaciones de cuarto orden, pero también se puede aplicar a ecuaciones de otros órdenes y tipos.

2. El último de los métodos que examinaremos se denomina método de diferencias finitas o método de redes.

La derivada de la función u con respecto a la variable x se define como el límite del cociente

$$\frac{u(x + \Delta x) - u(x)}{\Delta x}.$$

Este cociente se puede representar a su vez en la forma

$$\frac{1}{\Delta x} \int_x^{x+\Delta x} \frac{\partial u}{\partial x_1} dx_1,$$

y utilizando el bien conocido teorema del valor medio (capítulo 2, §8), tenemos:

$$\frac{u(x + \Delta x) - u(x)}{\Delta x} = \left. \frac{\partial u}{\partial x} \right|_{x=\xi},$$

donde ξ es un punto del intervalo

$$x < \xi < x + \Delta x.$$

Todas las derivadas segundas de u —tanto las derivadas mixtas como las derivadas con respecto a una variable— se pueden representar aproximadamente en forma de un cociente de diferencias. Así, pues, el cociente de diferencias

$$\frac{u(x + \Delta x) - 2u(x) + u(x - \Delta x))}{(\Delta x)^2}$$

viene representado en la forma

$$\begin{aligned} \frac{1}{\Delta x} \left[\frac{u(x + \Delta x) - u(x)}{\Delta x} - \frac{u(x) - u(x - \Delta x)}{\Delta x} \right] &= \\ &= \frac{1}{\Delta x} \left\{ \left[\frac{u(x_1 + \Delta x) - u(x_1)}{\Delta x} \right] \right|_{x_1=x-\Delta x}^{x_1=x} \right\}. \end{aligned}$$

Por el teorema del valor medio, el cociente de diferencias de la función

$$\phi(x_1) = \frac{u(x_1 + \Delta x) - u(x_1)}{\Delta x}$$

puede reemplazarse por el valor de la derivada. En consecuencia,

$$\frac{\phi(x_1) - \phi(x_1 - \Delta x)}{\Delta x} = \phi'(\xi),$$

donde ξ es un valor intermedio en el intervalo

$$x - \Delta x < \xi < x.$$

Por tanto,

$$\begin{aligned} \left(\frac{1}{\Delta x}\right)^2 [u(x + \Delta x) - 2u(x) + u(x - \Delta x)] &= \\ &= \frac{1}{\Delta x} [\phi(x) - \phi(x - \Delta x)] = \phi'(\xi). \end{aligned}$$

Por otra parte

$$\phi(\xi) = \frac{u(\xi + \Delta x) - u(\xi)}{\Delta x},$$

lo cual significa que

$$\phi'(\xi) = \frac{u'(\xi + \Delta x) - u'(\xi)}{\Delta x}.$$

Usando una vez más la fórmula de los incrementos finitos, vemos que

$$\phi'(\xi) = u''(\eta),$$

donde

$$\xi < \eta < \xi + \Delta x.$$

En consecuencia,

$$\left(\frac{1}{\Delta x}\right)^2 [u(x + \Delta x) - 2u(x) + u(x - \Delta x)] = u''(\eta),$$

donde $x - \Delta x < \eta < x + \Delta x$.

Si la derivada $u''(x)$ es continua y el valor de Δx es suficientemente pequeño, entonces $u''(\eta)$ apenas diferirá de $u''(x)$. Por tanto nuestra segunda derivada es arbitrariamente próxima al cociente de diferencias en cuestión. Exactamente de la misma forma podemos probar, por ejemplo, que la derivada segunda mixta

$$\frac{\partial^2 u}{\partial x \partial y}$$

puede ser representada aproximadamente por la fórmula

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{1}{\Delta x \Delta y} [u(x + \Delta x, y + \Delta y) - u(x + \Delta x, y) - u(x, y + \Delta y) + u(x, y)].$$

Volvamos ahora a nuestra ecuación en derivadas parciales.

Para precisar, supongamos que se trata de la ecuación de Laplace con dos variables independientes

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Además, supongamos que nos dan la función incógnita u sobre la frontera S del dominio Ω . Como aproximación supongamos que

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} &= \frac{u(x + \Delta x, y) - 2u(x, y) + u(x - \Delta x, y)}{(\Delta x)^2}, \\ \frac{\partial^2 u}{\partial y^2} &= \frac{u(x, y + \Delta y) - 2u(x, y) + u(x, y - \Delta y)}{(\Delta y)^2}. \end{aligned}$$

Si ponemos $\Delta x = \Delta y = h$, entonces

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} &= \frac{1}{h^2} [u(x + h, y) + u(x, y + h) + u(x - h, y) + \\ &\quad + u(x, y - h) - 4u(x, y)]. \end{aligned}$$

Recubramos ahora el dominio Ω mediante una red cuadriculada con vértices en los puntos $x = kh$, $y = bh$ (Fig. 4) y reemplacemos el dominio Ω por el polígono formado por aquellos cuadrados que caen en el interior de Ω , de modo que el contorno del dominio quede sustituido por una línea quebrada. Tomemos como valores de la función incógnita sobre esta línea quebrada los valores correspondientes al contorno de S . La ecuación de Laplace está aproximada entonces por la ecuación

$$u(x + h, y) + u(x, y + h) + u(x - h, y) + u(x, y - h) - 4u(x, y) = 0$$

para todos los puntos interiores al dominio. Esta ecuación se puede volver a escribir en la forma

$$u(x, y) = \frac{1}{4} [u(x + h, y) + u(x, y + h) + u(x - h, y) + u(x, y - h)].$$

Entonces el valor de u en un punto de la red, por ejemplo el punto 1 de la figura 4, es igual a la media aritmética de sus valores en los cuatro puntos adyacentes.

Supongamos que en el interior del polígono hay N puntos de nuestra red. En cada uno de tales puntos tendremos una ecuación. De esta manera obtenemos un sistema de N ecuaciones algebraicas con N incógnitas, cuya solución nos da los valores aproximados de la función u sobre el dominio Ω .

Puede demostrarse que para la ecuación de Laplace la solución se puede hallar con cualquier grado de exactitud.

El método de diferencias finitas reduce el problema a la resolución de un sistema de N ecuaciones con N incógnitas, donde éstas son los valores de la función deseada en los nudos de una red.

Además, puede demostrarse que el método de diferencias finitas es aplicable a otros problemas de física matemática:

otras ecuaciones diferenciales e integrales. Sin embargo, su aplicación en muchos casos implica un gran número de dificultades.

Puede resultar que la solución del sistema de N ecuaciones algebraicas con N incógnitas, construido por el método de redes, o no exista en general o dé un resultado que sea totalmente diferente del verdadero. Esto sucede cuando la resolución del sistema de ecuaciones conduce a acumulación de errores; cuanto más pequeña sea la longitud de los lados de los cuadrados de la red, mayor será el número de ecuaciones obtenidas, de modo que el error acumulado puede hacerse más grande.

En el ejemplo dado anteriormente de la ecuación de Laplace, esto no ocurre. Los errores al resolver el sistema no se acumulan, sino que, por el contrario, disminuyen sensiblemente si resolvemos el sistema, por ejemplo, por un método de aproximaciones sucesivas. Para la ecuación del flujo de calor y para la ecuación de ondas es

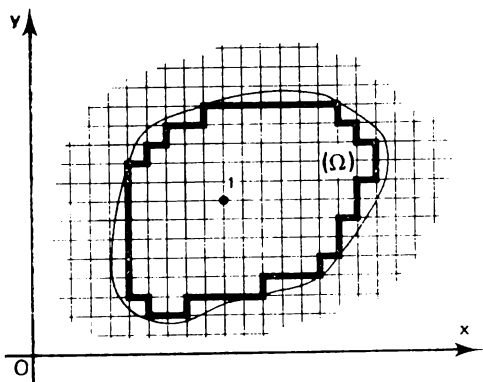


FIG. 4

esencial elegir las redes convenientemente. Para estas ecuaciones podemos obtener tanto buenos como malos resultados.

Si vamos a resolver alguna de estas ecuaciones por el método de redes, después de elegir una red para los valores de t no debemos tomar una demasiado fina para las variables espaciales, pues de otra manera obtendremos un sistema de ecuaciones sumamente insatisfactorio para los valores de la función incógnita; su resolución da un resultado que oscila rápidamente con amplitudes muy grandes y que está, por tanto, muy lejos del verdadero.

La gran variedad de resultados posibles se puede ver mejor con un sencillo ejemplo numérico. Consideremos la ecuación

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

para el flujo de calor en el caso en que la temperatura no dependa de y ni de z . Tomemos k como anchura de la malla de la red según los valores de t , y h como anchura según los valores de x :

$$\frac{\partial u}{\partial t} \approx \frac{u(t+k, x) - u(t, x)}{k},$$

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u(t, x+h) - 2u(t, x) + u(t, x-h)}{h^2}.$$

Entonces nuestra ecuación puede ser escrita aproximadamente en la forma

$$u(t+k, x) = \frac{k}{h^2} u(t, x+h) + \left(1 - 2 \frac{k}{h^2}\right) u(t, x) + \frac{k}{h^2} u(t, x-h).$$

Si para un valor de t correspondiente a un punto de la malla conocemos los valores de u en los puntos $x-h$, x y $x+h$, es fácil hallar el valor de u en el punto x y en el siguiente punto de la malla $t+k$. Supongamos que la constante k , es decir, la anchura de la malla con respecto a t , está ya elegida. Consideremos dos casos para la elección de h . Pongamos $h^2 = k$ en el primer caso y $h^2 = 2k$ en el segundo, y resolvamos el siguiente problema por el método de redes.

En el instante inicial, $u = 0$ para todos los valores negativos de x , y $u = 1$ para todos los valores no negativos de x . Escribiendo en

una línea los valores de la función incógnita u para el instante dado, tendremos dos tablas:

TABLA 1

$t \backslash x$	$-5h$	$-4h$	$-3h$	$-2h$	$-h$	0	h	$2h$	$3h$	$4h$	$5h$
0	0	0	0	0	0	1	1	1	1	1	1
k	0	0	0	0	1	0	1	1	1	1	1
$2k$	0	0	0	1	-1	2	0	1	1	1	1
$3k$	0	0	1	-2	4	-3	3	0	1	1	1
$4k$	0	1	-3	7	-9	10	-6	4	0	1	1
$5k$	1	-4	11	-19	26	-25	20	-10	5	0	1

TABLA 2

$t \backslash x$	$-5h$	$-4h$	$-3h$	$-2h$	$-h$	0	h	$2h$	$3h$	$4h$	$5h$
0	0	0	0	0	0	1	1	1	1	1	1
k	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1	1	1
$2k$	0	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	1	1	1	1
$3k$	0	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{7}{8}$	$\frac{7}{8}$	1	1	1
$4k$	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	$\frac{15}{16}$	1	1
$5k$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{13}{16}$	$\frac{13}{16}$	$\frac{31}{32}$	$\frac{31}{32}$	1

En la tabla 2 obtenemos valores que, para un instante dado, varían suavemente de un punto a otro. Esta tabla da una buena aproximación a la solución de la ecuación de flujo de calor. Por otra parte, en la tabla 1, en la cual, aparentemente, la exactitud habría sido

aumentada debido a una más fina división del intervalo de las x , los valores de u oscilan muy rápidamente de positivos a negativos y alcanzan cotas mucho mayores que las impuestas inicialmente. Está claro que en esta tabla los valores se hallan extraordinariamente lejos de la solución verdadera.

De estos ejemplos salta a la vista que si queremos utilizar el método de redes para obtener resultados dignos de confianza y suficientemente exactos, debemos proceder con gran prudencia en la elección de intervalos en la red e investigar previamente la aplicabilidad del método.

Las soluciones que se obtienen utilizando las ecuaciones de física matemática para estos u otros problemas de las ciencias nos dan una descripción matemática del curso o carácter esperado de los sucesos físicos descritos por estas ecuaciones.

Puesto que la construcción de un modelo se lleva a cabo mediante las ecuaciones de la física matemática, nos vemos forzados a ignorar, en nuestras abstracciones, muchos aspectos de estos acontecimientos, a desechar algunos como no esenciales y a seleccionar otros como básicos, de lo que se deduce que los resultados obtenidos no son absolutamente verdaderos. Lo son para ese esquema o modelo que hemos considerado, pero deben comprobarse siempre con experimentos para asegurarnos de que nuestro modelo del fenómeno es fiel al fenómeno real y de que lo refleja con un grado suficiente de exactitud.

El criterio último de la veracidad de los resultados es, por tanto, la experiencia práctica solamente. En último término no hay más que un criterio: la experiencia práctica, aunque ésta solamente puede ser propiamente comprendida a la luz de una profunda y bien desarrollada teoría.

En el caso de la cuerda vibrante de un instrumento musical, sólo podremos entender cómo produce sus tonos si estamos familiarizados con las leyes de superposición de oscilaciones características. Las relaciones que existen entre las frecuencias sólo pueden comprenderse si estudiamos cómo vienen determinadas aquellas por el material, por la tensión de la cuerda y por la manera de fijar los extremos. En este caso la teoría no se limita a proporcionar un método de calcular ciertas cantidades numéricas deseadas, sino que también indica precisamente cuáles de estas cantidades son de importancia fundamental, cómo tiene lugar el proceso físico exactamente y qué sería observado en él.

De esta forma, el dominio de la ciencia denominado física mate-

mática nació de las exigencias de la práctica, pero ejerció a su vez también cierta influencia sobre esa práctica y señaló caminos para ulteriores progresos.

La física matemática está muy estrechamente conectada con otras ramas del análisis matemático. No podemos, sin embargo, discutir estas conexiones aquí, ya que nos llevaría demasiado lejos.

§6. *Soluciones generalizadas*

La gama de problemas en los cuales un proceso físico viene descrito mediante funciones continuas y diferenciables puede ser ampliada de manera esencial introduciendo en la discusión soluciones discontinuas de estas ecuaciones.

En muchos casos está claro desde el principio que el problema en consideración no puede tener soluciones que sean dos veces continuamente diferenciables; en otras palabras, desde el punto de vista del enunciado clásico del problema dado en la sección anterior, tal problema no tiene solución. No obstante, el correspondiente proceso físico sí que ocurre, si bien en la indicada clase de funciones dos veces diferenciables no existen funciones que lo describan. Consideremos algunos ejemplos sencillos.

1. Si una cuerda consiste en dos trozos de diferente densidad, entonces en la ecuación

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (24)$$

el coeficiente será igual a una constante diferente en cada uno de los correspondientes trozos y, por tanto, la ecuación (24) no tendrá, en general, soluciones clásicas (dos veces continuamente diferenciables).

2. Supongamos que el coeficiente a es constante, pero que en la posición inicial la cuerda tiene la forma de una línea quebrada dada por la ecuación $u|_{t=0} = \phi(x)$. En el vértice de la línea quebrada, la función $\phi(x)$ evidentemente no puede tener derivada primera. Se puede probar que no existe ninguna solución clásica de la ecuación (24) que verifique las condiciones iniciales

$$u|_{t=0} = \phi(x), \quad u_t|_{t=0} = 0$$

(aquí y en lo que sigue, u_t denota $\partial u / \partial t$).

3. Dando un golpe brusco a cualquier segmento pequeño de la cuerda, las oscilaciones que resultan están descritas por la ecuación

$$\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} = f(x, t),$$

donde $f(x, t)$ corresponde al efecto producido y es una función discontinua, distinta de cero solamente sobre el pequeño segmento de cuerda y durante un corto intervalo de tiempo. Tal ecuación (como puede probarse con facilidad) tampoco puede tener soluciones clásicas.

Estos ejemplos muestran que exigiendo la continuidad de las derivadas para la solución deseada se restringe fuertemente la clase de problemas que podemos resolver. La búsqueda de una clase más amplia de problemas resolubles discurrió en primer lugar en el sentido de permitir discontinuidades de primera especie en las derivadas de orden máximo correspondientes a las funciones que sirven como soluciones del problema, funciones que deben verificar las ecuaciones excepto en los puntos de discontinuidad. Ocurre que las soluciones de una ecuación del tipo $\Delta u = 0$ o $\partial u / \partial t - \Delta u = 0$ no pueden tener tales discontinuidades (también llamadas débiles) en el interior del dominio de definición. Las soluciones de la ecuación de ondas no pueden tener discontinuidades débiles en las variables espaciales x, y, z y en t más que sobre superficies de una forma especial, llamadas superficies características. Si una solución $u(x, y, z, t)$ de la ecuación de ondas se considera como una función que define, para $t = t_1$, un campo escalar en el espacio (x, y, z) en el instante t_1 , entonces las superficies de discontinuidad para las derivadas segundas de $u(x, y, z, t)$ se moverán a través del espacio (x, y, z) con una velocidad igual a la raíz cuadrada del coeficiente de la laplaciana en la ecuación de ondas.

El segundo ejemplo referente a la cuerda muestra que también es necesario considerar soluciones en las cuales puedan existir primeras derivadas discontinuas; y en el caso de ondas sonoras y luminosas, debemos considerar incluso soluciones que sean ellas mismas discontinuas.

La primera cuestión que surge al investigar la introducción de soluciones discontinuas consiste en aclarar exactamente qué funciones discontinuas se pueden considerar como soluciones físicamente admisibles de una ecuación o del correspondiente problema físico. Podríamos, por ejemplo, suponer que una función arbitraria constante a trozos es una «solución simple» de la ecuación de Laplace o de la ecuación

de ondas, ya que verifica la ecuación fuera de las líneas de discontinuidad.

Para aclarar esta cuestión, lo primero que se debe garantizar es que en esa clase más amplia de funciones a la cual deben pertenecer las soluciones admisibles exista un teorema de unicidad. Es evidente que si, por ejemplo, admitimos funciones arbitrarias regulares a trozos, entonces esta exigencia no se satisfará.

Históricamente, el primer principio de selección de funciones admisibles fue que éstas debían ser el límite (en un sentido o en otro) de soluciones clásicas de la misma ecuación. Así, pues, en el ejemplo 2 una solución de la ecuación (24) correspondiente a la función $\phi(x)$, que no tiene derivada en un punto angular, puede hallarse como el límite uniforme de soluciones clásicas $u_n(x, t)$ de la misma ecuación correspondientes a las condiciones iniciales $u_n|_{t=0} = \phi_n(x)$, $u_{n,t}|_{t=0} = 0$, donde las $\phi_n(x)$ son funciones dos veces continuamente diferenciables que convergen uniformemente a $\phi(x)$ para $n \rightarrow \infty$.

En lo que sigue, en vez de este principio adoptaremos el siguiente: una función admisible u debe verificar, en vez de la ecuación $Lu = f$, una identidad integral que contenga una función arbitraria Φ .

Esta identidad se halla como sigue: multiplicamos ambos miembros de la ecuación $Lu = f$ por una función arbitraria Φ que tenga derivadas continuas —con respecto a todos sus argumentos— hasta el orden de la ecuación y que sea nula fuera del dominio finito D en el que está definida la ecuación. La ecuación así obtenida se integra sobre D y se transforma, integrando por partes, hasta que no contenga derivadas de u . Como resultado obtenemos la identidad deseada. Para la ecuación (24), por ejemplo, obtenemos la forma

$$\iint_D u \left[\frac{\partial^2 \Phi}{\partial t^2} - \frac{\partial^2 (a^2 \Phi)}{\partial x^2} \right] dx dt = 0.$$

S. L. Sobolev ha demostrado que para ecuaciones con coeficientes constantes estos dos principios de selección de soluciones admisibles (o, como se denominan normalmente ahora, generalizadas), son equivalentes. Pero, para ecuaciones con coeficientes variables, el primer principio puede resultar inaplicable, ya que estas ecuaciones pueden no tener en general soluciones clásicas (cf. ejemplo 1). El segundo de estos principios proporciona la posibilidad de seleccionar soluciones generalizadas con hipótesis muy amplias sobre las

propiedades de diferenciabilidad de los coeficientes de las ecuaciones. Es cierto que a primera vista este principio parece ser demasiado formal y tener un carácter puramente matemático, que no indica directamente cómo deben ser formulados los problemas de una forma análoga a los problemas clásicos.

Daremos aquí un método modificado que nos parece más apropiado físicamente, ya que está conectado directamente con el bien conocido principio de Hamilton.

Como ya es sabido, el análisis de los métodos de deducción de ecuaciones diversas de la física matemática llevó, en la primera mitad del siglo XIX, al descubrimiento de una nueva ley, conocida como principio de Hamilton. Partiendo de este principio, fue posible obtener de una manera uniforme todas las ecuaciones conocidas de la física matemática. Ilustraremos esto con el ejemplo del problema considerado en §3 para las oscilaciones de una cuerda de longitud finita con extremos fijos.

En primer lugar construyamos la llamada función de Lagrange (o lagrangiana) $L(t)$ para nuestra cuerda, la diferencia entre la energía cinética y la potencial. De lo dicho en §3 se sigue que

$$L(t) = \int_0^l \left(\frac{1}{2} \rho u_t^2 - \frac{T}{2} u_x^2 \right) dx.$$

De acuerdo con el principio de Hamilton, la integral

$$S = \int_{t_1}^{t_2} L(t) dt$$

toma su valor mínimo para la función $u(x, t)$ que corresponde al movimiento real de la cuerda, en contraposición con cualquier otra función $v(x, t)$ que sea igual a cero para $x = 0$ y $x = l$ y coincida con $u(x, t_1)$ y $u(x, t_2)$ para $t = t_1$ y $t = t_2$. Aquí t_1 y t_2 están fijados arbitrariamente, y las funciones v deben tener integrales finitas S . Como resultado de este principio, la llamada primera variación de S (cf. capítulo 8) debe ser igual a cero:

$$\delta S = \int_{t_1}^{t_2} \int_0^l (\rho u_t \Phi_t - T u_x \Phi_x) dx dt = 0, \quad (25)$$

donde $\Phi(x, t)$ es una función arbitraria diferenciable con respecto a x y t e igual a cero sobre los bordes del rectángulo $0 \leq x \leq l$, $t_1 \leq t \leq t_2$.

La ecuación (25) es también la condición que debe cumplir la función deseada $u(x, t)$. Si sabemos que $u(x, t)$ tiene derivadas de segundo orden, entonces la condición (25) puede escribirse en una forma diferente. Integrando (25) por partes y aplicando el lema fundamental del cálculo de variaciones, hallamos que $u(x, t)$ debe verificar la ecuación

$$\frac{\partial}{\partial t} \left(\varrho \frac{\partial u}{\partial t} \right) - \frac{\partial}{\partial x} \left(T \frac{\partial u}{\partial x} \right) = 0, \quad (26)$$

que es idéntica a (24), si ϱ y T son constantes y $T/\varrho = a^2$.

No es difícil ver que cualquier solución $u(x, t)$ de la ecuación (26) verifica la identidad (25) para toda Φ dada. El recíproco puede ser falso, ya que $u(x, t)$ puede no tener segundas derivadas en general. Así, pues, trabajando con la ecuación (25) en lugar de la (26), lo que hacemos es ampliar la gama de problemas resolubles.

Para determinar una oscilación específica de la cuerda, debemos añadir a las condiciones de contorno

$$u(0, t) = u(l, t) = 0 \quad (27)$$

las condiciones iniciales

$$\begin{aligned} u(x, 0) &= \phi_0(x), \\ u_t(x, 0) &= \phi_1(x). \end{aligned} \quad (28)$$

Si la solución se busca dentro de la clase de funciones continuamente diferenciables, entonces las condiciones (27) y (28) pueden desligarse de (25) como requisitos a verificar por separado. Pero si aceptamos que la solución propuesta sea «peor», entonces estas condiciones pierden su significado en la forma dada y deben ser parcial o totalmente incluidas en la identidad integral (25).

Por ejemplo, sea $u(x, t)$ continua para $0 \leq x \leq l$, $0 \leq t \leq T$, pero tal que sus derivadas primeras tengan discontinuidades. La segunda de las ecuaciones (28) pierde entonces su significado como condición limitadora. En este caso el problema puede ser enunciado como sigue: hallar una función continua u que cumpla la condición (27) y la primera de las (28) y para la cual la ecuación

$$\int_0^T \int_0^l (\varrho u_t \Phi_t - T u_x \Phi_x) dx dt + \int_0^l \phi_1 \Phi(x, 0) dx = 0 \quad (29)$$

sea idénticamente satisfecha por toda función continua $\Phi(x, t)$ igual a cero para $x = 0$, $x = l$ y $t = T$. Aquí, ambas funciones u y Φ deben tener derivadas primeras cuyos cuadrados sean integrables en el sentido de Lebesgue sobre el rectángulo $0 \leq x \leq l$, $0 \leq t \leq T$. Esta última exigencia para u significa que el valor medio con respecto al tiempo de la energía total de la cuerda

$$\frac{1}{2T} \int_0^T \int_0^l (\rho u_t^2 + T u_x^2) dx dt$$

debe ser finito. Tal restricción sobre la función u , y también sobre sus posibles variaciones Φ , es un resultado natural del principio de Hamilton.

La identidad (29) es precisamente la condición de que la primera variación del funcional

$$\tilde{S} = \int_0^T \int_0^l \left(\frac{\rho}{2} u_t^2 - \frac{T}{2} u_x^2 \right) dx dt + \int_0^l \phi_1 u|_{t=0} dx$$

sea igual a cero. Así, pues, el problema de la vibración de una cuerda fija en el caso considerado se puede reducir al de hallar el mínimo del funcional $\tilde{S}$ para todas las funciones $v(x, t)$ que son continuas, verifican la condición (27) y son iguales a $u(x, T)$ para $t = T$. Además, la función deseada debe verificar la primera de las condiciones (28).

Esta modificación del principio de Hamilton nos permite, no sólo ampliar la clase de soluciones admisibles de la ecuación (24), sino también enunciar un problema de contorno bien definido para ella.

El hecho de que estas soluciones generalizadas o alguna de sus derivadas no estén definidas en todos los puntos del espacio no conduce a ninguna contradicción con la experiencia, como señaló repetidamente N. M. Gjunter, cuyas investigaciones contribuyeron fundamentalmente al establecimiento de una nueva visión del concepto de solución de una ecuación de la física matemática.

Por ejemplo, si queremos determinar el flujo de un líquido en un conducto, en la presentación clásica debemos calcular el vector velocidad y la presión en todo punto de la corriente. Sin embargo, en la práctica nunca se plantea el calcular la presión en un punto, sino más bien la presión sobre una cierta área; ni tampoco interesa el vector velocidad en un punto dado, sino la cantidad de líquido que pasa a través de cierta área en la unidad de tiempo. Así, la definición

de solución generalizada propone en esencia el cálculo de sólo aquellas cantidades que tienen un significado físico inmediato.

Con el fin de poder resolver un mayor número de problemas, debemos buscar las soluciones en la clase más amplia posible de funciones para la que los teoremas de unicidad aún se verifican. Con frecuencia tal clase viene determinada por la naturaleza física del problema. De este modo, en mecánica cuántica no es la función de estado $\psi(x)$, definida como una solución de la ecuación de Schrödinger, la que tiene significado físico, sino más bien la integral $a_\nu = \int_E \psi(x) \psi_\nu(x) dx$, donde las ψ_ν son ciertas funciones para las que $\int_E \psi_\nu^2 dx < \infty$. Así, pues, la solución ψ debe ser buscada, no entre las funciones dos veces continuamente diferenciables, sino entre las de cuadrado integrable. En los problemas de electrodinámica cuántica no está aún establecido entre qué clases de funciones debemos buscar soluciones para las ecuaciones consideradas en esa teoría.

Los progresos de la física matemática durante los últimos treinta años tienen mucho que ver con esta nueva formulación de los problemas y con la creación del aparato matemático necesario para su resolución. Una de las facetas centrales de este aparato es el llamado teorema de inclusión de S. L. Sobolev.

Métodos particularmente convenientes de cálculo de soluciones en una u otra de estas clases de funciones son: el método de diferencias finitas, los métodos directos del cálculo de variaciones (método de Ritz y método de Trefftz), el de Galerkin, y los métodos de operadores funcionales. Estos últimos dependen básicamente de un estudio de transformaciones originadas por estos problemas. En §5 hemos hablado ya del método de diferencias finitas y del de Galerkin. Aquí expondremos las ideas fundamentales de los métodos directos del cálculo de variaciones.

Consideremos el problema de definir la posición de una membrana uniformemente estirada y de contorno fijo. Por el principio de energía potencial mínima en una posición de equilibrio estable, la función $u(x, y)$ debe dar el menor valor de la integral

$$J(u) = \iint_D (u_x^2 + u_y^2) dx dy$$

en comparación con todas las demás funciones continuamente diferenciables $v(x, y)$ que verifican la misma condición sobre el contorno, $v|_S = \phi$, que la función u . Con algunas restricciones sobre ϕ y sobre

el contorno S se puede probar que tal mínimo existe y que es alcanzado con una función armónica, de modo que la función deseada u es una solución del problema de Dirichlet $\Delta u = 0$, $u|_S = \phi$. El recíproco también es cierto: la solución del problema de Dirichlet da un mínimo para la integral J con respecto a todas las v que verifiquen la condición de contorno.

La demostración de la existencia de la función u , para la cual J alcanza su mínimo, y su cálculo con cualquier grado de exactitud puede hacerse, por ejemplo, de la siguiente manera (método de Ritz). Tomemos una familia infinita de funciones dos veces continuamente diferenciables $\{v_n(x, y)\}$, $n = 0, 1, 2, \dots$, iguales a cero sobre el contorno para $n > 0$, e iguales a ϕ para $n = 0$. Consideremos J para funciones de la forma

$$v = \sum_{k=1}^n C_k v_k + v_0,$$

donde n es fijo y los C_k son números arbitrarios. Entonces $J(v)$ será un polinomio de segundo grado en las n variables independientes $C_1, C_2, \dots, C_n$. Determinamos los C_k a partir de la condición de que este polinomio se haga mínimo. Esto conduce a un sistema de n ecuaciones algebraicas lineales con n incógnitas, cuyo determinante es distinto de cero. De este modo los números C_k están unívocamente determinados. Denotemos la correspondiente v por $v^n(x, y)$. Se puede probar que, si el sistema $\{v_n\}$ verifica una cierta condición de «completitud», las funciones v^n convergerán, cuando $n \rightarrow \infty$, a una función que será la solución deseada del problema.

Observemos finalmente que en este capítulo sólo hemos dado una descripción de los problemas lineales más sencillos de la mecánica y hemos ignorado muchas otras cuestiones (lejos aún de estar completamente resueltas) relacionadas con ecuaciones en derivadas parciales más generales.

Capítulo 7

CURVAS Y SUPERFICIES

§1. *Problemas y métodos en la teoría de curvas y superficies*

En los cursos de enseñanza escolar, la geometría sólo abarca las curvas más sencillas: líneas rectas, líneas quebradas, circunferencias y arcos de círculo; y en cuanto a superficies, únicamente planos, superficies de poliedros, esferas, conos y cilindros. En cursos más extensos se consideran otras curvas, principalmente las secciones cónicas: elipses, parábolas e hipérbolas. Pero el estudio de una curva o superficie arbitraria es completamente ajeno a la geometría elemental. A primera vista no está ni siquiera claro cuáles podrían ser las propiedades generales que hay que someter a investigación cuando hablamos de curvas y superficies arbitrarias. No obstante, tal investigación es completamente natural y necesaria.

En toda clase de actividad práctica y experiencia de la naturaleza encontramos constantemente curvas y superficies de muy diferentes formas. La trayectoria de un planeta en el espacio, de un barco en el mar, de un proyectil en el aire; la huella de un cincel sobre el metal, de una rueda sobre la carretera, de una pluma sobre una cinta; la forma de un árbol de levas que gobierna las válvulas de un motor; el perfil de un dibujo artístico; la forma de una cuerda colgante,

de un muelle espiral enrollado con algún propósito específico: ejemplos como éstos son interminables. Las superficies de diversos objetos, cáscaras delgadas, cisternas, el armazón de un aeroplano, envolturas, materiales laminados, proporcionan una diversidad interminable de superficies. Los métodos para la elaboración de productos, las propiedades ópticas de distintos objetos, las líneas de corriente de los cuerpos, la rigidez o deformabilidad de láminas delgadas, estas y otras muchas características dependen en gran parte de la forma geométrica de las superficies de los objetos.

Por supuesto, la estría hecha por un cincel sobre el metal no es una curva matemática. Una cisterna, aún de paredes delgadas, no es una superficie matemática. No obstante, en una primera aproximación —que para el estudio de muchas cuestiones es suficiente—, los objetos reales se pueden representar matemáticamente por curvas y superficies.

Al introducir el concepto de curva matemática, nos olvidamos de todas las razones por las cuales es imposible disminuir sin límite un espesor. Mediante este concepto abstracto conseguimos representar aquellas propiedades (completamente concretas) de un objeto que no desaparecen al disminuir su espesor y anchura en relación con su longitud.

Análogamente, si hacemos caso omiso de las limitaciones de nuestra capacidad para disminuir el espesor de una lámina o para determinar con exactitud los contornos reales de un objeto dado, llegamos al concepto de superficie matemática. No daremos una descripción rigurosa de estos conceptos bien conocidos, sino solamente haremos notar que las definiciones matemáticas exactas no son sencillas y pertenecen a la topología.

Finalmente, una importante fuente de interés por diversas curvas y superficies ha sido el desarrollo del análisis matemático. Basta recordar, por ejemplo, que una curva es la representación geométrica de una función, que es el concepto más importante del análisis. Por otra parte, todo el mundo está familiarizado con las gráficas, aun cuando no lo esté con el estudio del análisis.

En la geometría elemental, tal como fue creada por los antiguos griegos, no había nada sobre curvas y superficies arbitrarias. Sin embargo, desde la geometría analítica elemental estamos acostumbrados a decir «cada curva está representada por una ecuación» o «toda ecuación en dos variables x e y representa una curva en el plano coordenado». Análogamente, las coordenadas de una superficie están dadas por las ecuaciones $z = f(x, y)$ o $F(x, y, z) = 0$, y en

general el método coordenado, al establecer una estrecha conexión entre la geometría elemental y el análisis, nos permite definir muchas curvas y superficies diferentes.

Pero la geometría analítica, restringida como está a los métodos del álgebra y la geometría elemental, no sirve más que para el estudio de algunos tipos específicos de figuras. El estudio de curvas y superficies arbitrarias representa una nueva rama de las matemáticas, conocida como *geometría diferencial*.

Admitamos de entrada que la geometría diferencial impone sobre las curvas y superficies ciertas condiciones, que derivan de los métodos del análisis. Sin embargo, esto no es una limitación esencial sobre la gama de curvas y superficies permisibles, ya que en la gran mayoría de los casos aquéllas son capaces de representar objetos reales con el grado necesario de precisión. El nombre «geometría diferencial» indica ya los métodos a que recurre la teoría; su herramienta básica es el cálculo diferencial, e investiga primordialmente las propiedades «diferenciales» de las curvas y superficies, es decir, sus propiedades «en un punto» o «locales»¹. De este modo, la dirección de una curva en un punto está determinada por su tangente en ese punto, y la medida en que se «retuerce» está descrita por su curvatura (la definición exacta de este término se dará más adelante). La geometría diferencial investiga las propiedades de pequeños segmentos de curvas y superficies, y sólo en una etapa más avanzada procede al estudio de sus propiedades «globales».

El desarrollo de la geometría diferencial está inseparablemente conectado con el desarrollo del análisis. Las operaciones básicas del análisis —diferenciación e integración— tienen un significado geométrico inmediato. Como mencionamos en el capítulo 2, diferenciar una función $f(x)$ corresponde a trazar la tangente a la curva

$$y = f(x).$$

La pendiente de la recta tangente (es decir, la tangente trigonométrica del ángulo que forma con el eje Ox) es precisamente la derivada

¹ Las propiedades de curvas y superficies «en un punto» son aquellas propiedades que sólo dependen de un entorno arbitrariamente pequeño del punto. Las propiedades de este tipo se definen en términos de las derivadas (en el punto dado) de las funciones que aparecen en las ecuaciones de la curva o superficie. Por esta razón, la geometría diferencial impone condiciones que garantizan la aplicabilidad del cálculo diferencial, exigiendo que la curva o superficie esté definida por funciones con un número suficiente de derivadas.

$f'(x)$ de la función $f(x)$ en el punto correspondiente (Fig. 1), y el área «bajo la curva»

$$y = f(x)$$

es precisamente la integral $\int_a^b f(x) dx$ de esta función, calculada entre los límites correspondientes. Igual que en análisis investigamos funciones arbitrarias, en geometría diferencial examinamos curvas y superficies arbitrarias. En análisis, el primer objeto de estudio es el

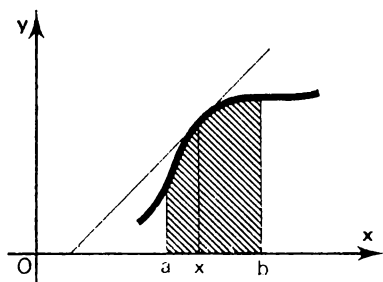


FIG. 1

recorrido general de una curva en el plano, sus elevaciones y depresiones, su mayor o menor curvatura, la dirección de su convexidad, sus puntos de inflexión, etc. La estrecha conexión entre el análisis y las curvas viene indicada por el título del primer libro de texto de análisis, escrito por el matemático francés l'Hôpital en 1665: «Análisis infinitesimal aplicado al estudio de curvas.»

Hacia mediados del siglo XVIII, el cálculo diferencial e integral estaba ya suficientemente des-

arrollado (gracias a los inmediatos sucesores de Newton y Leibnitz) para permitir una más profunda aplicación a la geometría. En realidad, es sólo desde este momento cuando se puede hablar propiamente de una teoría de curvas y superficies. Para superficies, y para curvas en el espacio, los problemas análogos son inmensamente más ricos en contenido que para curvas planas; con el paso del tiempo, estos problemas rebasaron el marco de una simple aplicación del análisis a la geometría y condujeron a la formación de una teoría independiente. Durante la segunda mitad del siglo XVIII, numerosos matemáticos participaron en la construcción de los elementos de esta teoría: Clairaut, Euler, Monge y otros, entre los cuales Euler debe ser considerado como el fundador de la teoría general de superficies. El primer trabajo amplio sobre curvas y superficies fue el libro de Monge «Aplicación del análisis a la geometría», publicado en 1795².

² Gaspard Monge (1746-1828) no sólo fue un científico destacado, sino también un activo revolucionario francés (ministro de asuntos navales, y después director de la fábrica de cañones y pólvora). Siguiendo un camino característico de la burguesía francesa de la época, pasó de jacobino a partidario del emperador Napoleón.

A la vista de las investigaciones de estos matemáticos, y en particular del libro de Monge, podemos comprender con facilidad el interés que surgió por la geometría diferencial. Este surgimiento fue debido a las demandas de la mecánica, la física y la astronomía, es decir, en última instancia, a las necesidades de la tecnología y la industria, para las cuales los resultados de la geometría elemental eran completamente insuficientes.

El trabajo clásico de Gauss (1777-1855) en la teoría de superficies está también relacionado con cuestiones prácticas. Su libro «Investigaciones generales sobre superficies curvas», publicado en 1827, es básico para la geometría diferencial de superficies como rama independiente de las matemáticas. Sus métodos generales y problemas, que discutiremos más adelante en §4, se originaron sobre todo en las necesidades prácticas de la confección de mapas. El problema de la cartografía consiste en hallar una representación lo más exacta posible de partes de la superficie de la tierra sobre un plano. Una representación completamente exacta es imposible, pues las relaciones mutuas de las distintas longitudes están necesariamente distorsionadas debido a la curvatura de la tierra. Así, pues, se tiene el problema de encontrar los métodos más exactos posibles. El trazado de mapas se remonta a la Antigüedad, pero la creación de una teoría general es un logro de tiempos recientes y no habría sido posible sin la teoría general de superficies y los métodos generales de análisis matemático. Observemos que uno de los problemas matemáticos difíciles de la cartografía fue investigado por P. L. Chebichev (1821-1894), quien obtuvo importantes resultados relativos a redes de curvas sobre superficies. Sus investigaciones también surgieron de problemas puramente prácticos.

Las cuestiones generales de la deformación de una superficie y su proyección sobre otra constituyen una de las principales ramas de la geometría. Importantes resultados en esta dirección fueron obtenidos en 1838 por F. Minding (1806-1885), profesor de la Universidad de Dorpat (ahora Tartu).

Hacia la segunda mitad del siglo pasado, la teoría de curvas y superficies quedó ya establecida en sus rasgos básicos, siempre que nos refiramos a la «geometría diferencial clásica» (en contraste con las más modernas direcciones discutidas más tarde en §5). Las ecuaciones fundamentales de la teoría de curvas, las llamadas fórmulas de Frenet, habían sido ya obtenidas, y en 1853 K. M. Peterson (1828-1881), alumno de Minding en la Universidad de Tartu, descubrió e investigó las ecuaciones básicas de la teoría de superficies, ecuaciones

que fueron redescubiertas y publicadas quince años más tarde por el matemático italiano Codazzi, con cuyo nombre están usualmente asociadas. Después de graduarse por la Universidad en Tartu, Peterson vivió y trabajó en Moscú como profesor de instituto. Aunque nunca ocupó una posición académica acorde con sus notables logros científicos, fue sin embargo uno de los fundadores de la Sociedad Matemática de Moscú y de la revista «Matematičeski Sbornik», publicada en Moscú desde 1866 hasta el momento presente. La escuela de Moscú de geometría diferencial empieza con Peterson.

Los resultados de la geometría diferencial «clásica» fueron compendiados por el geómetra francés Darboux en su obra en cuatro tomos «Lecciones sobre la teoría general de superficies», publicada entre 1887 y 1896. En el presente siglo se sigue estudiando la geometría diferencial clásica; pero el foco de interés en lo relativo a curvas y superficies se ha desplazado hacia otras direcciones en las que las clases de figuras estudiadas son cada vez más amplias.

§2. La teoría de curvas

Diversos métodos de definición de curvas en geometría diferencial

Del análisis y la geometría analítica estamos acostumbrados a la idea de definir curvas por medio de ecuaciones. En un sistema de coordenadas rectangulares en el plano, una curva puede estar dada por la ecuación

$$y = f(x),$$

o por la ecuación más general

$$F(x, y) = 0.$$

Sin embargo, este método de definición es apropiado solamente para una curva plana, es decir, una línea en el plano. Necesitamos también un método de escribir ecuaciones de curvas en el espacio no situadas en un plano. Un ejemplo de este tipo de curvas es la hélice (Fig. 2).

Para los propósitos de la geometría diferencial, y para muchas otras cuestiones también, resulta muy conveniente representar una curva como la traza del movimiento continuo de un punto. Naturalmente, la curva dada puede estar originada de una forma completamente distinta, pero siempre podemos imaginarla como el camino de un punto que se mueve a lo largo de ella.

Supongamos que hemos fijado un sistema de coordenadas cartesianas en el espacio. Si un punto móvil X describe una curva desde el instante $t = a$ a $t = b$, entonces las coordenadas de este punto móvil vienen dadas por las funciones del tiempo $x(t)$, $y(t)$ y $z(t)$; el vuelo de un aeroplano o un proyectil son ejemplos. Recíprocamente, dadas inicialmente las funciones $x(t)$, $y(t)$ y $z(t)$, podemos servirnos de ellas para definir las coordenadas de un punto móvil X que describa alguna curva. En consecuencia, las curvas en el espacio pueden venir dadas por tres ecuaciones de la forma

$$x = x(t), \quad y = y(t), \quad z = z(t).$$

Análogamente, una curva plana está definida por dos ecuaciones:

$$x = x(t), \quad y = y(t).$$

Esta es la manera más general de definir curvas.

Como ejemplo consideremos la hélice. Está originada por el movimiento espiral de un punto que gira uniformemente alrededor de una recta (el eje de la hélice) y al mismo tiempo se desplaza uniformemente en una dirección paralela a este eje. Tomeámos como eje de la hélice el Oz y supongamos que en el instante $t = 0$ el punto está sobre Ox . Queremos hallar ahora cómo dependen sus coordenadas del tiempo. Si el movimiento paralelo al eje Oz tiene velocidad c , entonces evidentemente la distancia recorrida en esta dirección en el tiempo t será

$$z = ct,$$

Por otra parte, si ϕ es el ángulo de rotación alrededor del eje Oz y a es la distancia del punto a este eje, entonces, como se puede ver en la figura 2,

$$x = a \cos \phi, \quad y = a \operatorname{sen} \phi.$$

Puesto que la rotación es uniforme, el ángulo ϕ es proporcional al tiempo; esto es, $\phi = \omega t$, donde ω es la velocidad angular de la rotación.

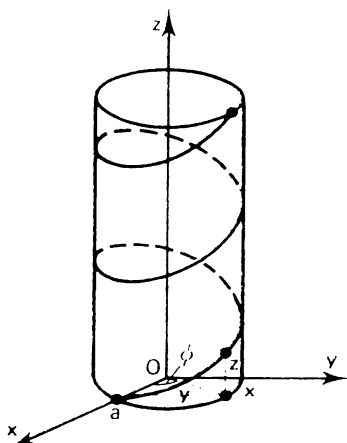


FIG. 2

De esta manera obtenemos

$$x = a \cos \omega t, \quad y = a \sin \omega t, \quad z = ct.$$

Estas son las ecuaciones de la hélice que describirá el punto móvil al variar t .

Naturalmente, la variable t o, como se denomina normalmente, el parámetro, no tiene que representar necesariamente el tiempo. El parámetro conocido t se puede sustituir por otro; podemos introducir, por ejemplo, un parámetro u mediante la fórmula $t = u^3$, o, en general, mediante $t = f(u)$ ³. En geometría, la elección más natural de parámetro es la longitud s del arco de la curva medido a partir de un punto fijo A sobre la curva. Cada valor posible de la longitud s representa un arco correspondiente AX . De este modo, la posición de X está totalmente determinada por el valor de s , y las coordenadas del punto X vendrán dadas por las funciones de la longitud de arco s

$$x = x(s), \quad y = y(s), \quad z = z(s).$$

Todas estas formas de definir curvas, así como otras posibles⁴, abren la posibilidad de un cálculo numérico. Solamente cuando las curvas han sido definidas por ecuaciones pueden estudiarse sus propiedades mediante el análisis matemático.

En la geometría diferencial de curvas planas existen tres conceptos básicos: longitud, tangente y curvatura. Para las curvas en el espacio, existen además el plano osculador y la torsión. Procedemos ahora a exponer el sentido y significado de estos conceptos.

Longitud

Todo el mundo tiene en la mente una idea natural de lo que se entiende por longitud. Sin embargo, esta idea es preciso convertirla en una definición exacta de la longitud de una curva matemática, una definición con un carácter numérico específico que nos permita calcular la longitud de una curva con el grado deseado de exactitud y, en consecuencia, argumentar sobre longitudes de una manera rigurosa. La misma observación se aplica a todos los conceptos matemá-

³ Aquí, estrictamente hablando, es necesario que la función f sea monótona.

⁴ Una curva en el espacio puede venir dada también como intersección de dos superficies, definidas por las ecuaciones: $F(x, y, z) = 0$, $G(x, y, z) = 0$, es decir, la curva viene dada por este par de ecuaciones. En discusiones de carácter teórico, las curvas se suelen representar por un vector variable: la posición del punto X de la curva está determinada por el vector $\vec{r} = \overrightarrow{OX}$ que va del origen a este punto. Al variar el vector $\vec{r}$, su extremo x recorre la curva dada (Fig. 3).

ticos. La transición de ideas no formales a medidas exactas y definiciones representa el paso del conocimiento precientífico de los objetos a una teoría científica. La necesidad de una definición precisa de longitud surgió, en última instancia, de las exigencias de la tecnología y las ciencias naturales, cuyo desarrollo requería la investigación de las propiedades de longitudes, áreas y otras entidades geométricas.

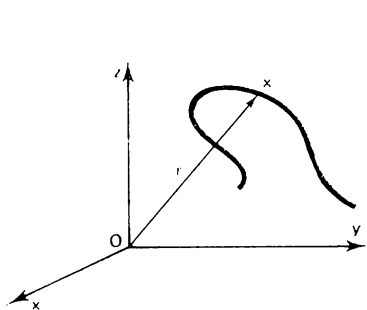


FIG. 3

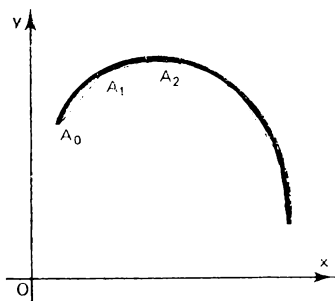


FIG. 4

Una sencilla y muy útil definición de longitud es la siguiente: la *longitud* de una curva es el límite de la longitud de poligonales inscritas en ella con la condición de que sus vértices se aproximen cada vez más unos a otros sobre la misma.

Esta definición surge naturalmente de los métodos cotidianos de medida. Tomemos una sucesión de puntos sobre la curva, $A_0, A_1, A_2, \dots$ (Fig. 4), y midamos las distancias entre ellos. La suma de estas distancias (que será la longitud de la poligonal) expresará aproximadamente la longitud de la curva. Para definir la longitud más exactamente, es natural tomar los puntos A más próximos entre sí, de tal forma que la poligonal se adapte a la curva más estrechamente. El valor exacto de la longitud se define como el límite de estas aproximaciones cuando los puntos A se eligen arbitrariamente próximos entre sí ⁵.

⁵ La existencia del límite indicado, es decir, la longitud de la curva, no es evidente en principio, incluso para curvas contenidas en un dominio acotado. Si la curva es muy retorcida, su longitud puede ser muy grande; matemáticamente es posible construir una curva plana tan «retorcida» que ninguno de sus arcos tenga longitud finita debido a que las longitudes de las poligonales inscritas sobrepasan cualquier cota.

De este modo la definición anterior de longitud es una generalización —basada en dar pasos cada vez más finos— de una manera completamente práctica de medir longitudes.

De esta definición es fácil deducir la fórmula para calcular longitudes cuando la curva viene dada analíticamente. Observemos, sin embargo, que las fórmulas matemáticas son útiles para algo más que para hacer simples cálculos: son un breve enunciado de teoremas que establecen conexiones entre diferentes entes matemáticos. El signi-

ficado teórico de tales conexiones puede superar con mucho el valor «computacional» de la fórmula. Por ejemplo, la importancia del teorema de Pitágoras, expresado mediante la fórmula

$$c^2 = a^2 + b^2,$$

no se limita al cálculo del cuadrado de la hipotenusa c , sino

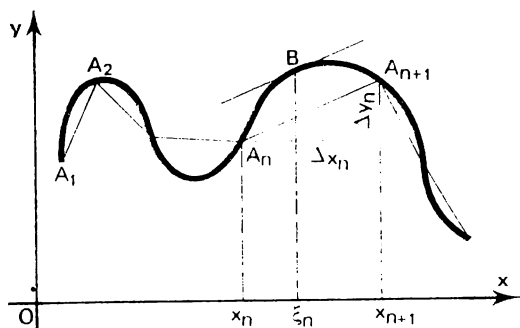


FIG. 5

que estriba sobre todo en el hecho de que expresa una relación entre los lados de un triángulo rectángulo.

Introduzcamos ahora una fórmula para la longitud de una curva plana, dada en coordenadas cartesianas mediante la ecuación $y = f(x)$, suponiendo que la función $f(x)$ tenga derivada primera.

Inscribimos una poligonal en la curva (Fig. 5). Sean A_n, A_{n+1} dos de sus vértices adyacentes de coordenadas x_n, y_n y x_{n+1}, y_{n+1} . El segmento $A_n A_{n+1}$ es la hipotenusa de un triángulo rectángulo cuyos catetos son iguales a

$$\Delta x_n = |x_{n+1} - x_n|, \quad \Delta y_n = |y_{n+1} - y_n|.$$

Así pues, por el teorema de Pitágoras,

$$A_n A_{n+1} = \sqrt{(\Delta x_n)^2 + (\Delta y_n)^2} = \sqrt{1 + \left(\frac{\Delta y_n}{\Delta x_n}\right)^2} \Delta x_n.$$

Es fácil ver que, si la recta que une los puntos A_n y A_{n+1} se traslada paralelamente a sí misma, entonces en el instante en que abandona a la curva asumirá la posición de una tangente a ésta en algún punto B , es decir, sobre el arco de curva $A_n A_{n+1}$ existe al menos un punto en el cual la tangente tiene la misma dirección que la cuerda $A_n A_{n+1}$. (Esta conclusión obvia puede deducirse fácilmente con una demostración rigurosa.)

De este modo podemos sustituir la razón $\Delta y_n / \Delta x_n$ por la pendiente de la tangente en B , o lo que es lo mismo, por la derivada $y'(\xi_n)$, donde ξ_n es la abscisa del punto B . Ahora la longitud de un lado de la poligonal viene expresada por

$$\overline{A_n A_{n+1}} = \sqrt{1 + y'^2(\xi_n)} \Delta x_n.$$

La longitud total de la poligonal es la suma de las longitudes de sus lados. Denotando la suma con el símbolo Σ , tenemos

$$S_n = \sum \sqrt{1 + y'^2(\xi_n)} \Delta x_n.$$

Para obtener la longitud de la curva, debemos pasar al límite con la condición de que el mayor de los valores Δx_n tienda a cero,

$$s = \lim_{\Delta x \rightarrow 0} \sum \sqrt{1 + y'^2(\xi_n)} \Delta x_n.$$

Pero este límite es exactamente la integral definida en el capítulo 2, la integral de la función $\sqrt{1 + y'^2}$. De este modo la longitud de una curva plana está expresada por la fórmula

$$s = \int_a^b \sqrt{1 + y'^2} dx, \quad (1)$$

donde los límites de integración a y b son los valores de x en los extremos del arco de curva.

La correspondiente fórmula para la longitud de una curva en el espacio se deduce esencialmente de la misma forma.

El cálculo efectivo de la longitud por medio de estas fórmulas no es siempre sencillo, naturalmente. Así, el cálculo de la circunferencia de un círculo a partir de la fórmula (1) es bastante complicado. Sin embargo, como hemos dicho ya, el interés de las fórmulas no

se reduce al cálculo; en particular (1) es también importante para estudiar las propiedades generales de la longitud, sus relaciones con otros conceptos, etc. En el capítulo 8 tendremos ocasión de utilizar la fórmula (1).

Tangente

La tangente a una curva plana fue ya considerada en el capítulo 2. Su sentido para una curva en el espacio es completamente análogo. Para definir la tangente en un punto A , tomemos un punto X sobre

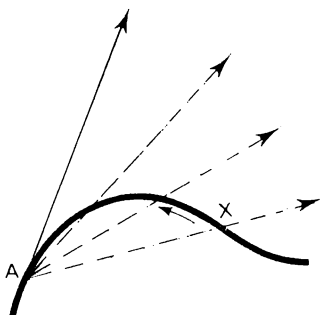


FIG. 6

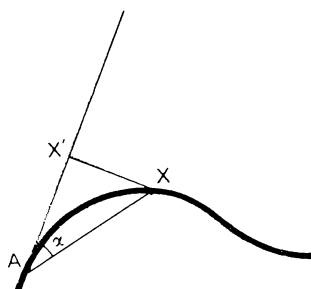


FIG. 7

la curva, distinto de A , y consideremos la secante AX . Dejemos que X se aproxime a A a lo largo de la curva. Si la secante AX converge a alguna posición límite, entonces la recta en esta posición límite se denomina tangente en el punto A ⁶.

Si distinguimos entre origen y final de la curva y, como consecuencia de ello, establecemos el orden en que son recorridos los puntos de aquélla, entonces podemos saber cuál de los puntos A y X está antes y cuál después. (Por ejemplo, si un tren viaja de Moscú a Vladivostok, entonces Omsk precede evidentemente a Irkutsk.) Así podemos definir un sentido, a lo largo de la secante, del primer punto al segundo. El límite de tales «secantes orientadas» nos da una «tangente orientada». En la figura 6, la flecha muestra el sentido en que se atraviesa el punto A . Para el movimiento de un punto a lo

⁶ La posición límite de la secante puede no existir, como se puede ver en el ejemplo de la figura 13, capítulo 2. La curva representada por $y = x \operatorname{sen} 1/x$ oscila en las proximidades de cero de tal forma que la secante OA , cuando A se aproxima a O , oscila constantemente entre las rectas OM y OL .

largo de la curva, la velocidad en cada instante está dirigida según la tangente.

La tangente tiene una importante propiedad geométrica: cerca del punto de tangencia la curva se desvía menos —en un sentido bien definido— de esta recta que de cualquier otra. En otras palabras, la distancia de los puntos de la curva a la tangente es muy pequeña en comparación con su distancia al punto de tangencia. O más precisamente: la razón XX'/AX (Fig. 7) tiende a cero cuando X se aproxima a A ⁷. De este modo, si sustituimos un pequeño arco de curva por un segmento de la tangente, el error que se comete es pequeño en comparación con la longitud del segmento. Este procedimiento nos permite a menudo simplificar las demostraciones, ya que en el paso al límite da resultados completamente exactos.

Es interesante observar que para una curva que no sea una línea recta, es decir, que no tenga una dirección en el sentido elemental, hemos logrado definir (asociando la curva con una recta) su dirección en cada punto. De este modo hemos ampliado el concepto de dirección, dándole un significado que no tenía previamente. Este nuevo concepto de dirección refleja la naturaleza real del movimiento a lo largo de una curva; en cada instante el punto se mueve en alguna dirección definida, que varía continuamente.

Curvatura

Para poder juzgar si un camino, una varilla delgada o una línea de un dibujo está más o menos curvado no es necesario ser un matemático. Pero incluso para los problemas más simples de mecánica,

una mirada casual no es suficiente; necesitamos una exacta descripción cuantitativa de la curvatura. Esta se obtiene, dando expresión precisa a nuestra impresión intuitiva de la curvatura, como la rapidez de cambio de dirección de la curva.

Sea A un punto de la curva y M un punto próximo a A (Fig. 8).

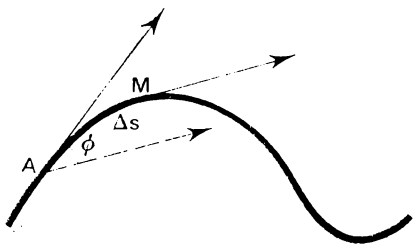


FIG. 8

⁷ Este resultado se deduce inmediatamente de la misma definición de tangente. Evidentemente, como se muestra en la figura 7, $XX'/AX = \sin \alpha$, donde α es el ángulo entre la tangente y la secante AX . Así, cuando $\alpha \rightarrow 0$, XX'/AX también tiende a cero.

El ángulo entre las tangentes en estos dos puntos expresa lo que la curva ha variado de dirección en el arco de A a M . Denotemos este ángulo por ϕ . La variación media de dirección (más precisamente, la variación media por unidad de longitud del camino a lo largo del arco AM de longitud Δs) será evidentemente $\phi/\Delta s$. Entonces la curvatura, es decir, la variación de dirección de la curva en el punto A , está naturalmente definida por la razón $\phi/\Delta s$ cuando $M \rightarrow A$; dicho

con otras palabras, cuando $\Delta s \rightarrow 0$. Así, pues, la curvatura viene definida por la fórmula

$$k = \lim_{\Delta s \rightarrow 0} \frac{\phi}{\Delta s}$$

Como ejemplo particular, consideremos la curvatura de la circunferencia de un círculo (Fig. 9). Evidentemente, el ángulo ϕ entre los radios OA y OM es igual al ángulo ϕ entre las

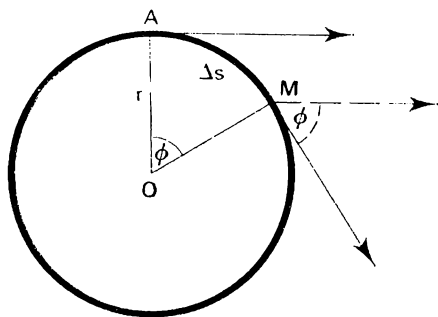


FIG. 9

tangentes en los puntos A y M , ya que las tangentes son perpendiculares a los radios. El arco AM , que subtende el ángulo ϕ , tiene de longitud $\Delta s = \phi r$, así que

$$\frac{\phi}{\Delta s} = \frac{1}{r}.$$

Esto significa que la razón $\phi/\Delta s$ es constante y que la curvatura de la circunferencia, como valor límite de esta razón, es igual en todos los puntos a la inversa del radio ⁸.

Deduzcamos la fórmula de la curvatura de una curva plana dada por la ecuación $y = f(x)$. Como punto inicial para la longitud de arco tomemos uno fijo N (Fig. 10). El ángulo ϕ entre las tangentes en los puntos A y M es evidentemente igual a la diferencia de los ángulos de inclinación de las tangentes en A y M .

$$\phi = |\Delta \alpha|.$$

⁸ Observemos que, en general, el concepto de curvatura de una curva en un punto se puede definir comparando la curva con la circunferencia de un cierto círculo, que representa el papel de un modelo o patrón para la curvatura. En efecto: la curvatura de una curva resulta ser igual al recíproco del radio del (único) círculo que se adapta más estrechamente a la curva en el entorno del punto.

Como el ángulo α puede ser decreciente, tomamos el valor absoluto $|\Delta\alpha|$. Nos interesa calcular el valor

$$k = \lim_{\Delta s \rightarrow 0} \frac{\phi}{\Delta s} = \lim_{\Delta s \rightarrow 0} \frac{|\Delta\alpha|}{\Delta s} = \lim_{\Delta x \rightarrow 0} \frac{\frac{|\Delta\alpha|}{\Delta x}}{\frac{\Delta s}{\Delta x}} = \frac{|\alpha'|}{s'}.$$

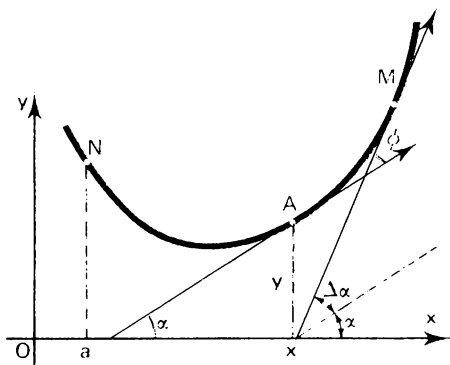


FIG. 10

La longitud del arco de la curva NA está expresada por la integral

$$s = \int_a^x \sqrt{1 + y'^2} dx,$$

de modo que

$$s' = \sqrt{1 + y'^2}.$$

Falta hallar α' . Sabemos que $\operatorname{tg} \alpha = y'$; así, pues, $\alpha = \operatorname{arc} \operatorname{tg} y'$. Derivando esta ecuación con respecto a x , obtenemos

$$\alpha' = \frac{1}{1 + y'^2} y''.$$

De este modo, tenemos finalmente

$$k = \frac{|\alpha'|}{s'} = \frac{|y''|}{(1 + y'^2)^{3/2}}.$$

Las correspondientes fórmulas para otros métodos de representación de curvas en el plano y en el espacio se dan en los cursos usuales de análisis o geometría diferencial.

Esta fórmula nos permite dar otra interpretación geométrica de la curvatura, útil en muchas cuestiones: la curvatura de una curva en un punto está dada por la fórmula

$$k = \lim_{l \rightarrow 0} \frac{2h}{l^2},$$

donde h es la distancia de un segundo punto de la curva a la tangente en el punto dado, y l es la longitud del segmento de tangente comprendido entre el punto de tangencia y la proyección del otro punto sobre ella (figura 11).

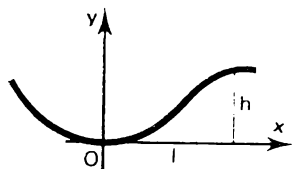


FIG. 11

Para probar esto elijamos un sistema de coordenadas rectangulares tal que el origen sea el punto dado de la curva, y el eje OX , la tangente en este punto (Fig. 11). (Por sencillez supongamos que la curva es plana.) Entonces $y' = 0$ y $k = |y''|$.

Desarrollando la función $y = f(x)$ por la fórmula de Taylor, obtenemos $y = \frac{1}{2} y'' x^2 + \epsilon x^2$ (donde hemos tenido en cuenta que $y' = 0$). Aquí $\epsilon \rightarrow 0$ cuando $x \rightarrow 0$. En consecuencia, deducimos que $k = |y''| = \lim_{x \rightarrow 0} 2|y|/x^2$, y por tanto, como $|y| = h$, $x^2 = l^2$, tenemos

$$k = \lim_{l \rightarrow 0} \frac{2h}{l^2}.$$

Esta fórmula muestra que la curvatura describe la rapidez con que la curva abandona la tangente.

Volvamos ahora a algunas aplicaciones muy importantes de la curvatura a problemas de la mecánica.

Consideremos primero el siguiente problema. Sea una cuerda flexible tendida sobre un soporte (Fig. 12) de tal forma que permanezca en un plano. Queremos hallar la presión de la cuerda sobre el soporte en cada punto, o para ser más exactos, definir el límite

$$p = \lim_{\Delta s \rightarrow 0} \frac{P}{\Delta s}, \quad (2)$$

donde P es la magnitud de la fuerza $\mathbf{P}$ que actúa sobre el soporte a lo largo de un trozo de longitud Δs que contiene al punto dado. Supongamos por simplicidad que la magnitud T de la tensión $\mathbf{T}$ es la misma en todos los puntos de la cuerda.

Consideremos ahora el punto A y el segmento de cuerda AB ⁹. Sobre este segmento AB de longitud Δs , además de la reacción del soporte, solamente actúan dos fuerzas externas, las tensiones en los extremos, que son iguales en magnitud y están dirigidas según las

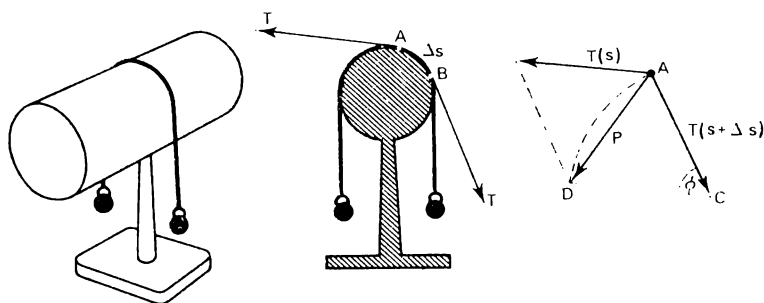


FIG. 12

tangentes en los extremos del segmento. De este modo, la fuerza $\mathbf{P}$ ejercida por la cuerda sobre el soporte es igual a la suma geométrica de las tensiones en los extremos. Tal como puede verse en la figura 12, el vector $\mathbf{P}$ es la base AD del triángulo isósceles CAD . Los dos lados iguales de este triángulo tienen longitud T , y el ángulo en el vértice C es igual al cambio de dirección de la tangente al pasar de A a B .

Al decrecer Δs , el ángulo ϕ decrece y el ángulo entre $\mathbf{P}$ y la tangente en el punto A se aproxima a un ángulo recto. Por tanto, la presión es perpendicular a la tangente.

Para hallar la magnitud de la presión, hagamos uso del hecho de que un arco pequeño de circunferencia tiene aproximadamente la misma longitud que la cuerda que lo subtiende. De este modo reem-

⁹ Sería más natural tomar un segmento con el punto A en su interior; esto no cambiaría el resultado, pero haría el cálculo algo más complicado.

plazamos la longitud de la cuerda AD , es decir, la magnitud P , por la longitud $T\phi$ del arco AD . Entonces, por la fórmula (2), obtenemos

$$p = \lim_{\Delta s \rightarrow 0} \frac{P}{\Delta s} = \lim_{\Delta s \rightarrow 0} \frac{T\phi}{\Delta s} = T \lim_{\Delta s \rightarrow 0} \frac{\phi}{\Delta s} = Tk.$$

Por consiguiente, la presión en cada punto es igual al producto de la curvatura por la tensión sobre la cuerda, y es ejercida perpendicularmente a la tangente en este punto.

Consideremos un segundo problema. Sea un punto matemático (un cuerpo muy pequeño) que se mueve a lo largo de una curva plana con una velocidad de magnitud constante v . ¿Cuál es la aceleración en un punto dado A ? Por definición, la aceleración es igual al límite de la razón del incremento de velocidad (durante el tiempo Δt) al incremento Δt de tiempo. En la velocidad interviene no sólo la magnitud, sino también la dirección, por lo cual habrá que considerar la variación en el vector velocidad. El problema matemático de hallar la magnitud de la aceleración consiste, pues, en hallar el límite

$$w = \lim_{\Delta t \rightarrow 0} \frac{|\mathbf{v}(t + \Delta t) - \mathbf{v}(t)|}{\Delta t},$$

donde $\mathbf{v}(t)$ es la velocidad en el punto A , y $|\mathbf{v}(t + \Delta t) - \mathbf{v}(t)|$ es la longitud del vector diferencia de velocidades. El límite que nos concierne puede también escribirse así:

$$\lim_{\Delta s \rightarrow 0} \frac{|\mathbf{v}(t) + \mathbf{v}(t + \Delta t)|}{\Delta s} \lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t},$$

donde Δs es la longitud del arco AB recorrido durante el tiempo Δt . Volviendo a la figura 13 y observando que la velocidad en cada punto está dirigida según la tangente, mientras permanece constante en magnitud, vemos geoméricamente que el cálculo de la suma $|\mathbf{v}(t) + \mathbf{v}(t + \Delta t)|$ es idéntico al cálculo del vector $\mathbf{P}$ en el problema precedente. Así, podemos aprovechar aquel resultado y, cambiando tensión por velocidad, escribir

$$\lim_{\Delta s \rightarrow 0} \frac{|\mathbf{v}(t) + \mathbf{v}(t + \Delta t)|}{\Delta s} = vk.$$

Por otra parte, $\lim_{\Delta t \rightarrow 0} \Delta s / \Delta t = v$. Llegamos así al resultado de que la aceleración de un cuerpo en movimiento uniforme a lo largo de

una curva es igual al producto de la curvatura por el cuadrado de la velocidad

$$w = kv^2, \quad (3)$$

y su dirección es la de la normal a la curva, es decir, la de la recta perpendicular a la tangente.

El haber recurrido aquí a una analogía geométrica, que nos ha permitido utilizar la solución del problema de la presión ejercida

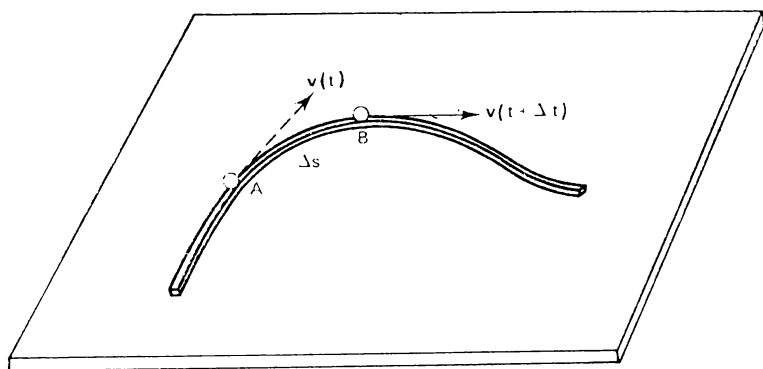


FIG. 13

por una cuerda para resolver el de la aceleración de una partícula, demuestra una vez más cuán útil es hacer abstracción de las propiedades concretas y particulares de un fenómeno y pasar a los correspondientes conceptos y resultados matemáticos, pues de este modo podemos hacer uso de estos resultados en las más diversas situaciones.

Observamos también que la curvatura, que desde un punto de vista mecánico refleja el cambio de dirección del movimiento, está estrechamente relacionada con las fuerzas que originan este cambio. La ecuación que expresa esta relación se deduce fácilmente multiplicando la ecuación (3) por la masa m del punto móvil. Tenemos

$$F_n = mw = v^2 mk.$$

Aquí F_n es la magnitud de la componente normal de la fuerza que actúa sobre el punto.

Plano osculador

Aunque una curva en el espacio no está contenida en un plano, es posible, en general, asociar a cada punto de la curva un plano P que en un entorno del punto está más próximo a la curva que cualquier otro plano. Dicho plano se denomina *plano osculador* de la curva en el punto.

Naturalmente, el plano osculador, como plano más próximo a la curva dada, pasa por el punto A y contiene la tangente T a la misma.

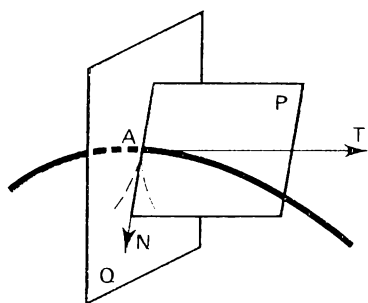


FIG. 14

Pero existen muchos otros planos que contienen el punto A y la recta T . Para elegir de entre ellos el que menos se desvía de la curva estudiamos la desviación de ésta a partir de la tangente. Para este propósito veamos cómo discurre la curva a lo largo de la tangente T ; en otras palabras, proyectemos nuestra curva sobre el *plano normal* Q , que es perpendicular a T en el punto A (Fig. 14). La proyección sobre el plano Q de un segmento de nuestra curva

que contenga a A forma una nueva curva —indicada en la figura 14 por una línea de puntos—, que suele tener una cúspide en el punto A . Si la curva así obtenida tiene una tangente N en A , entonces el plano P determinado por T y N será naturalmente el más próximo a la curva original en el entorno del punto A , es decir, será el plano osculador en el punto A . Se puede probar que cuando las funciones que definen la curva original tienen derivadas segundas y la curvatura en el punto A no es cero, entonces el plano osculador existe necesariamente, y su ecuación se puede expresar de modo muy sencillo en términos de las primeras y segundas derivadas de las funciones que definen la curva.

Vimos anteriormente que las propiedades de la tangente nos permiten considerar un segmento pequeño de una curva plana como si fuera recto, cometiendo así un error que es pequeño en comparación con la longitud del segmento; análogamente, las propiedades del plano osculador nos permiten considerar un pequeño segmento de una curva en el espacio como si fuera una curva plana —su proyección sobre el plano osculador—; aquí el error cometido será pequeño

en comparación con el cuadrado de la longitud del segmento de la curva.

Existen muchas rectas en el espacio que son perpendiculares a la tangente; todas ellas forman el plano normal en el punto dado de la curva. Entre estas rectas hay una, N , que está en el plano osculador y que se denomina la *normal principal* a la curva. Por lo común se suele fijar un sentido para ella: el de la concavidad de la proyección de la curva sobre el plano osculador. La normal principal juega el mismo papel para una curva en el espacio que la normal ordinaria (única) para una curva plana. En particular, si una cuerda delgada sometida a una tensión T se coloca en forma de una curva alabeada sobre un soporte, entonces la presión de la cuerda sobre éste tiene en cada punto la magnitud Tk y la dirección de la normal principal. Si un punto material se mueve a lo largo de una curva alabeada con una velocidad de magnitud constante v , entonces su aceleración es igual a kv^2 y tiene la dirección de la normal principal.

Torsión

De un punto a otro a lo largo de la curva la posición del plano osculador variará en general. Igual que la variación de dirección de la tangente caracterizaba a la curvatura, así la variación de dirección del plano osculador caracteriza una nueva cantidad, la *torsión* de la curva. Aquí, como en el caso de la curvatura, la variación se toma con respecto a la longitud de arco; esto es, si ψ es el ángulo entre los planos osculadores en un punto fijo A y en un punto próximo X , y si Δs es la longitud del arco AX , entonces la torsión τ en el punto A se define como el límite¹⁰

$$\tau = \lim_{\Delta s \rightarrow 0} \frac{\psi}{\Delta s}.$$

El signo de la torsión depende del lado de la curva hacia el que gira el plano osculador al moverse a lo largo de la curva.

Podemos imaginar el plano osculador como la pala de un ventilador sobre la que se hubiesen dibujado la tangente y la normal

¹⁰ Se puede demostrar que una hélice tiene la misma torsión en todos sus puntos y por consiguiente que podemos definir la torsión de una curva comparándola con la (única) hélice que mejor aproxima la curva en el entorno del punto dado. La torsión también caracteriza la forma en que una curva alabeada dada difiere de una curva plana. Con cierta analogía a la curvatura, caracteriza la velocidad con que la curva se separa de su plano osculador.

principal. En todo instante la tangente está girando en la dirección de la normal a una velocidad que viene determinada por la curvatura, mientras que el plano osculador rota alrededor de la tangente con una velocidad y dirección determinadas por la torsión.

Utilizando resultados muy simples de la teoría de ecuaciones diferenciales se puede probar un teorema fundamental que dice, toscamente hablando, que dos curvas con la misma curvatura y la misma torsión son idénticas una a la otra. Aclaremos esta idea. Si nos movemos a lo largo de la curva a distintas distancias de nuestro punto inicial A , llegaremos a puntos donde la curvatura k y la torsión τ tendrán distintos valores, dependientes de s . De este modo $k(s)$ y $\tau(s)$ serán ciertas funciones bien definidas de la longitud de arco s .

El teorema en cuestión dice que, si dos curvas tienen idénticas curvatura y torsión como funciones de la longitud de arco, entonces las curvas son idénticas (es decir, se puede mover rígidamente una de ellas hasta que coincida con la otra). De esta manera, la curvatura y la torsión como funciones de la longitud de arco definen completamente una curva salvo su posición en el espacio; describen todas las propiedades de la curva expresando las relaciones entre su longitud, su curvatura y su torsión. De esta forma, los tres conceptos constituyen una especie de fundamento esencial para cuestiones concernientes a curvas. Con su ayuda podemos también expresar los conceptos más simples de la teoría de superficies, de la cual pasamos a ocuparnos a continuación.

Naturalmente, la teoría de curvas no ha sido agotada por nuestras presentes observaciones. Hay muchos otros conceptos relativos a esta clase de figuras: tipos especiales de curvas, familias de curvas, posición de curvas sobre superficies, forma de una curva como un todo, etc. Estas cuestiones y los métodos de contestarlas están conectados con casi todas las ramas de las matemáticas. El tipo de problemas que se pueden resolver con la teoría de curvas es extremadamente rico y variado.

§3. *Conceptos básicos en la teoría de superficies*

Métodos fundamentales de definir una superficie

Si queremos estudiar superficies por medio del análisis debemos, por supuesto, definir las analíticamente. La manera más simple es mediante una ecuación

$$z = f(x, y),$$

en la cual x , y y z son coordenadas cartesianas de un punto que está sobre la superficie. Aquí la función $f(x, y)$ no tiene que estar definida necesariamente para todo x, y ; su dominio puede tener distintas formas. De este modo, la superficie de la figura 15 está dada por la función $f(x, y)$, definida en el interior de una corona circular. Ejemplos de superficies dadas por ecuaciones de la forma $z = f(x, y)$ son también corrientes en la geometría analítica. Sabemos, por ejemplo, que la ecuación $z = Ax + By + C$ representa un plano, y $z = x^2 + y^2$ un paraboloide de revolución (Fig. 16). Para aplicar el cálculo diferencial es necesario que la función $f(x, y)$ tenga de-

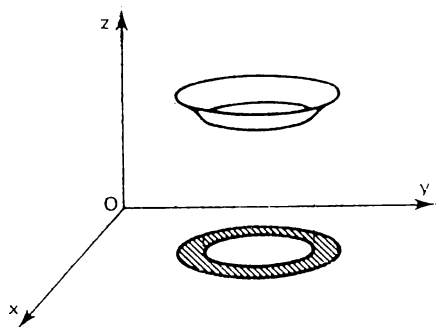


FIG. 15

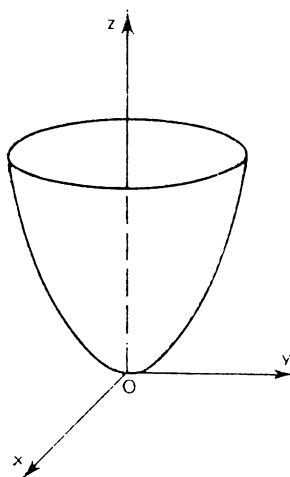


FIG. 16

derivadas primeras, segundas, y a veces también de orden superior. Una superficie dada por una ecuación de este tipo se llama *regular*. Geométricamente significa (aunque no muy exactamente) que la superficie se curva sin roturas u otras singularidades. Las superficies que no tienen esta propiedad, por ejemplo, las que tienen cúspides, roturas u otras singularidades, requieren un nuevo tipo de investigación (cf. §5).

Sin embargo, no toda superficie, aun sin singularidades, puede ser representada por una ecuación de la forma $z = f(x, y)$. Si cada par de valores de x, y en el dominio de $f(x, y)$ da una z completamente determinada, entonces toda recta paralela al eje Oz debe cortar a la superficie en no más de un punto (Fig. 17). Incluso superficies tan

simples como esferas o cilindros no pueden ser representadas globalmente por una ecuación de la forma $z = f(x, y)$. En estos casos, la superficie está definida de alguna otra manera, por ejemplo mediante una ecuación de la forma $F(x, y, z) = 0$. Así, una esfera de radio R con centro en el origen tiene la ecuación

$$x^2 + y^2 + z^2 = R^2.$$

La ecuación $x^2 + y^2 = r^2$ da un cilindro de radio r .

Cuando el estudio se reduce solamente a pequeños segmentos de la superficie —como es usualmente el caso en la geometría dife-

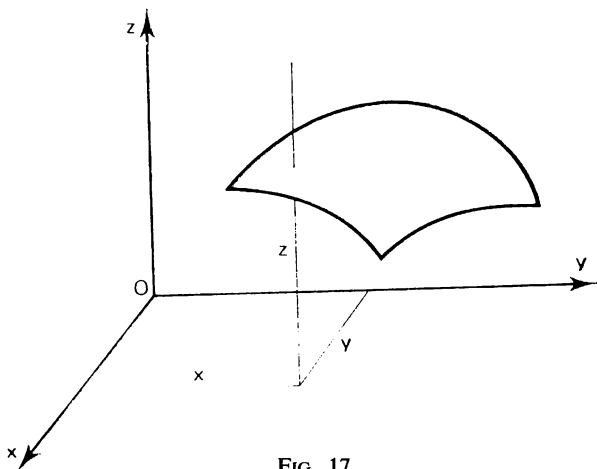


FIG. 17

rencial clásica— la definición de una superficie por una ecuación $z = f(x, y)$ es perfectamente general, ya que todo segmento suficientemente pequeño de una superficie regular puede ser representado en esta forma. Tomamos esta manera como básica, y dejamos otros métodos de definir superficies para considerarlos más tarde en las secciones 4 y 5.

Plano tangente

Así como una *curva* regular tiene en cada punto una tangente que está próxima a la curva en un entorno del punto, también las superficies pueden tener, en cada uno de sus puntos, un *plano tangente*.

La definición exacta es como sigue. Un plano P que pase por un punto M de una superficie F se dice que es tangente a la superficie F en este punto si el ángulo α entre el plano P y la secante MX , trazada desde M a un punto X de la superficie, converge a cero cuando X tiende a M (Fig. 18). Todas las tangentes a curvas que pasen por el punto M y estén sobre la superficie, evidentemente yacen en el plano tangente.

Una superficie F se llama regular si tiene un plano tangente en cada punto y si, al pasar de un punto a otro, la posición de este plano varía continuamente.

Cerca del punto de tangencia, la superficie se separa muy poco de su plano tangente: si el punto X se aproxima al M a lo largo de la superficie, entonces la distancia de X al plano tangente se hace más y más pequeña, incluso en comparación con su distancia al

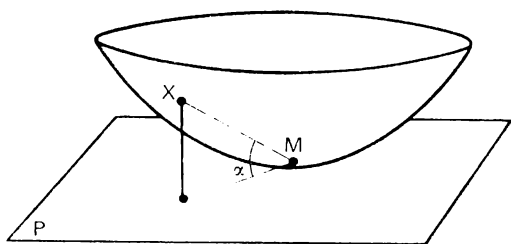


FIG. 18

punto M (el lector puede comprobar fácilmente esto considerando el modo en que X se aproxima a M en la figura 18). De este modo, la superficie en las proximidades del punto M puede decirse que se convierte gradualmente en el plano tangente. En una primera aproximación, un pequeño segmento o, como se suele llamar, un «elemento» de la superficie se puede reemplazar por un segmento del plano tangente. La perpendicular al plano tangente que pasa por el punto de tangencia viene a ser una perpendicular a la superficie en este punto y se denomina una *normal*.

La posibilidad de reemplazar un elemento de la superficie por un segmento del plano tangente es útil en muchas situaciones. Por ejemplo, la reflexión de la luz sobre una superficie curva tiene lugar de la misma forma que la reflexión sobre un plano, es decir, la dirección del rayo reflejado se define por la ley usual de la reflexión: el rayo incidente y el reflejado están en el mismo plano que la normal a la superficie y forman ángulos iguales con esta normal (Fig. 19); todo ocurre como si la reflexión tuviera lugar en el plano tangente. Análogamente, en la refracción de la luz por una superficie curva, cada

rayo es refractado por un elemento de la superficie siguiendo la ley usual de la refracción, como si el elemento fuera plano. Estos hechos son la base para todos los cálculos de reflexión y refracción de la luz en los aparatos ópticos. Además, por ejemplo, dos cuerpos sólidos en contacto tienen un plano tangente común en su punto de contacto. Este contacto se establece sobre un elemento de superficie, y la presión de un cuerpo sobre el otro, en ausencia de fricción, está dirigida según la normal en el punto de contacto. Esto también es cierto

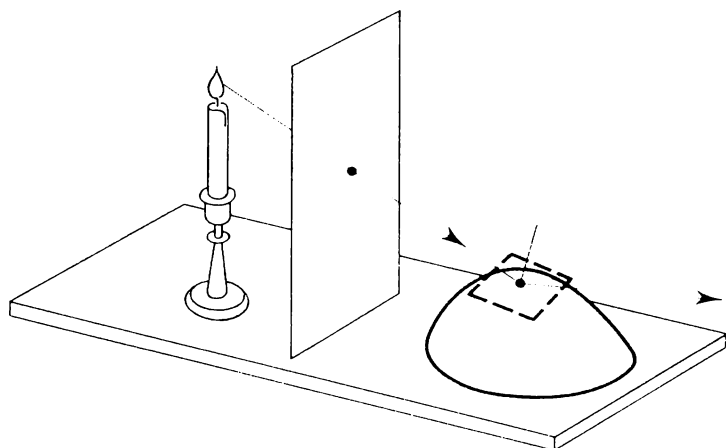


FIG. 19

cuando los cuerpos son tangentes en más de un punto, y en ese caso la presión está dirigida según las normales respectivas en cada punto de contacto.

La sustitución de elementos de una superficie por segmentos de los planos tangentes puede también servir como base para una definición del área de distintas superficies. La superficie se descompone en trozos pequeños $F_1, F_2, \dots, F_n$, y cada trozo se proyecta sobre un plano tangente a la superficie en algún punto de este trozo (Fig. 20). Obtenemos así varias regiones planas $P_1, P_2, \dots, P_n$, la suma de cuyas áreas da una aproximación del área de la superficie. Esta se define como el límite de las sumas de las áreas de los segmentos $P_1, P_2, \dots, P_n$, con la condición de que las particiones de la super-

ficie sean cada vez más finas ¹¹. De aquí podemos deducir una expresión exacta del área en forma de una integral doble.

Estas observaciones muestran claramente el significado del concepto de plano tangente. Sin embargo, en muchas cuestiones la representación aproximada de un elemento de una superficie por medio de un plano es inadecuada, siendo necesario considerar la curvatura de la superficie.

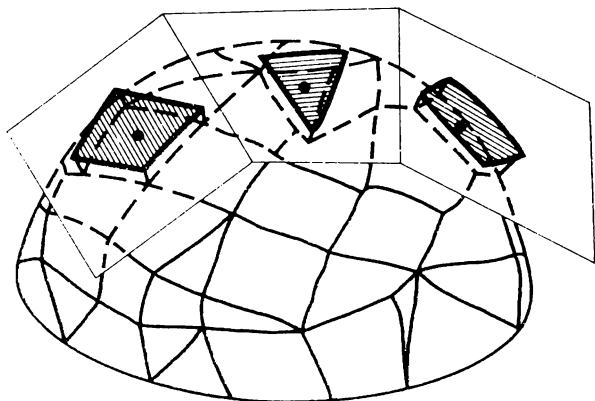


FIG. 20

Curvatura de curvas sobre una superficie

La curvatura de una superficie en un punto dado está caracterizada por la rapidez con que se separa de su plano tangente. Pero en direcciones diferentes, la superficie puede separarse de su plano tangente con distinta rapidez. Así, la superficie de la figura 21 se separa del plano P más rápidamente en la dirección OA que en la dirección OB . Por tanto, es natural definir la curvatura de una superficie en un punto por medio del conjunto de curvaturas de todas las curvas situadas sobre aquélla y que pasan por el punto dado en diferentes direcciones.

Esto se hace como sigue. Construimos el plano tangente P en el punto M y elegimos una orientación específica para la normal (Fig. 22). Luego consideramos aquellas curvas que son secciones de

¹¹ Esta es exactamente la expresión para el área que se utilizó en la sección 1 del capítulo 8.

la superficie al cortarla por planos que pasan por la normal en el punto M ; estas curvas se denominan *secciones normales*. La curvatura de una sección normal está afectada de un signo, que es «más» si la sección es cóncava en la dirección de la normal, y «menos» si es cóncava en el sentido opuesto. Así, en una superficie que tenga la

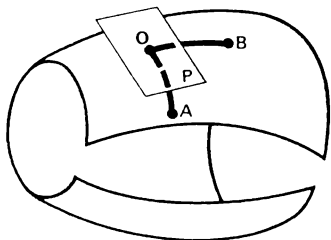


FIG. 21

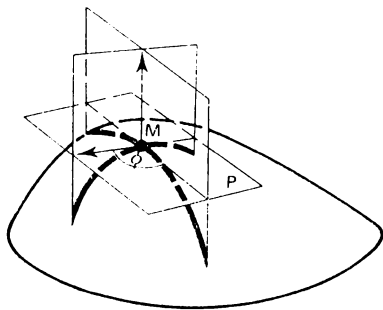


FIG. 22

forma de una silla de montar —como la de la figura 23, donde la flecha indica el sentido de la normal a la superficie— la curvatura de la sección MA es positiva, mientras que la de la sección MB es negativa.

Una sección normal viene determinada por el ángulo ϕ que forma su plano con algún rayo inicial situado en el plano tangente (Fig. 22). Si conocemos la curvatura de la sección normal $k(\phi)$ en función del ángulo ϕ , tendremos una imagen bastante completa del comportamiento de la superficie en las proximidades del punto M .

Una superficie puede estar curvada de muchas maneras diferentes, por lo que podría parecer que la dependencia de la curvatura k con respecto al ángulo ϕ es arbitraria. En realidad esto no es así. Para las superficies estudiadas en geometría diferencial existe una ley simple, debida a Euler, que establece la conexión entre las curvaturas de las secciones normales que pasan por un punto dado en diversas direcciones.

Se demuestra que en cada punto de una superficie hay dos direcciones particulares tales que

1. Son mutuamente perpendiculares.
2. Las curvaturas k_1 y k_2 de las secciones normales en estas

direcciones son los valores menor y mayor de las curvaturas de todas las secciones normales ¹².

3. La curvatura $k(\phi)$ de la sección normal que forma un ángulo ϕ con la sección de curvatura k_1 está expresada por la fórmula

$$k(\phi) = k_1 \cos^2 \phi + k_2 \sin^2 \phi. \quad (4)$$

Tales direcciones se denominan *direcciones principales*, y las curvaturas k_1 y k_2 se llaman *curvaturas principales de la superficie* en el punto dado.

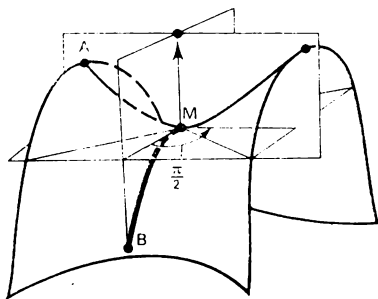


FIG. 23

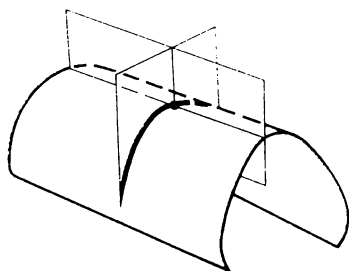


FIG. 24

Este teorema de Euler muestra que, a pesar de que existe una gran diversidad de superficies, la forma de éstas en el entorno de cada punto sólo puede ser una de entre un número muy limitado de tipos completamente definidos, con una exactitud del segundo orden de magnitud en comparación con la distancia al punto dado. En efecto, si k_1 y k_2 tienen el mismo signo, entonces el signo de $k(\phi)$ es constante y la superficie en las proximidades del punto tiene la forma que se ve en la figura 22. Si k_1 y k_2 tienen signos opuestos, por ejemplo $k_1 > 0$ y $k_2 < 0$, entonces la curvatura de la sección normal cambia de signo evidentemente. Esto se deriva del hecho de que para $\phi = 0$ la curvatura $k = k_1 > 0$, y para $\phi = \pi/2$ tenemos $k = k_2 < 0$.

A partir de la fórmula (4) para $k(\phi)$ no es difícil probar que

¹² En el caso particular $k_1 = k_2$ la curvatura de todas las secciones es la misma; esto ocurre, por ejemplo, en la esfera.

cuando ϕ varía de 0 a T el signo de $k(\phi)$ cambia dos veces ¹³, de modo que en las proximidades del punto la superficie tiene la forma de una silla de montar (Fig. 23).

Cuando uno de los números k_1 y k_2 es igual a cero, la curvatura tiene siempre el mismo signo, excepto para un valor de ϕ para el cual se anula. Esto ocurre, por ejemplo, para todo punto de un cilindro (Fig. 24). En el caso general, en las proximidades de tal punto la superficie tiene una forma parecida a la de un cilindro.

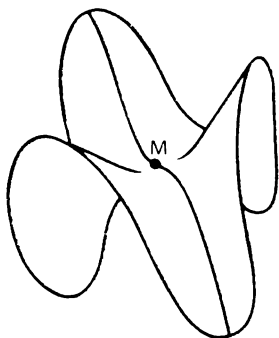


FIG. 25

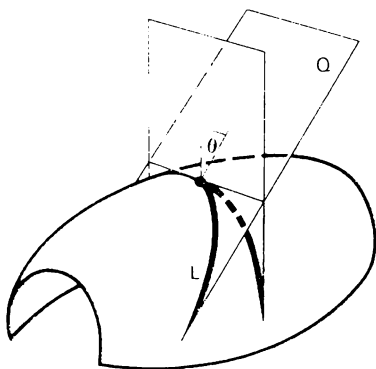


FIG. 26

Finalmente, para $k_1 = k_2 = 0$ todas las secciones normales tienen curvatura nula. En las proximidades de un punto de este tipo, la superficie está especialmente «próxima» a su plano tangente. Tales puntos se denominan *puntos planos*. Un ejemplo de este tipo de puntos se da en la figura 25 (el punto M). Las propiedades de una superficie en las proximidades de un punto plano pueden ser muy complicadas.

Consideremos ahora una sección de la superficie al cortar ésta por un plano arbitrario Q (Fig. 26) que no pase por la normal. La curvatura k_1 de una curva tal como L , como demostró Meusnier ¹⁴, está relacionada mediante una fórmula muy sencilla con la curva-

¹³ Es una cuestión muy sencilla probar que $k(\phi) = k_1 \cos^2 \phi + k_2 \sin^2 \phi$ se anula para $\phi = \arctg \sqrt{-k_1/k_2}$ y $\phi = \pi - \arctg \sqrt{-k_1/k_2}$, cambiando de signo la primera vez de más a menos, y la segunda, de menos a más.

¹⁴ Meusnier (1754-1793) fue un matemático francés, alumno de Monge; fue general del ejército revolucionario y murió de las heridas recibidas en el campo de batalla.

tura k_N de la sección normal en la misma dirección, es decir, la que corta al plano tangente en la misma recta. Esta conexión viene expresada mediante la fórmula

$$k_L = \frac{|k_N|}{\cos \theta},$$

donde θ es el ángulo entre la normal y el plano Q . La exactitud de esta fórmula se hace patente de una forma muy conveniente sobre una esfera.

Finalmente, se demuestra que la curvatura de *cualquier* curva que, perteneciendo a la superficie, tenga a Q como plano osculador es idéntica a la curvatura de la intersección de Q con la superficie.

De este modo, si conocemos k_1 y k_2 , la curvatura de cualquier curva sobre la superficie está determinada por la dirección de su tangente y el ángulo entre su plano osculador y la normal a la superficie. En consecuencia, el carácter de la curvatura de una superficie en un punto dado está determinado por los números k_1 y k_2 . Sus valores absolutos son iguales a las curvaturas de dos secciones normales perpendiculares entre sí, y sus signos indican la dirección de la concavidad de las respectivas secciones normales con respecto a una dirección convenida de la normal.

Demostraremos ahora los teoremas de Euler y Meusnier mencionados anteriormente.

1. Para la demostración del teorema de Euler necesitamos el siguiente lema. Si la función $f(x, y)$ tiene derivadas segundas continuas en un punto, entonces los ejes de coordenadas se pueden girar un ángulo α tal que en el nuevo sistema de coordenadas la derivada cruzada $f_{x'y'}$ sea igual a cero en este punto¹⁵. Recordemos que después de la rotación de ejes las nuevas variables x', y' están relacionadas con x e y mediante las fórmulas

$$x = x' \cos \alpha - y' \sin \alpha; \quad y = x' \sin \alpha + y' \cos \alpha$$

(capítulo 3, sección 7). Para la demostración del lema observemos que

$$\frac{\partial x}{\partial x'} = \cos \alpha, \quad \frac{\partial y}{\partial x'} = \sin \alpha, \quad \frac{\partial x}{\partial y'} = -\sin \alpha, \quad \frac{\partial y}{\partial y'} = \cos \alpha.$$

¹⁵ Denotaremos las derivadas parciales con subíndices; por ejemplo, en lugar de $\partial f / \partial x$ escribimos f_x , en lugar de $\partial^2 f / \partial y^2$ escribimos f_{yy} , etc.

Calculando la derivada $f_{x'y'}$ por la regla de la cadena, llegamos, después de simplificar, al resultado

$$f_{x'y'} = f_{xy} \cos 2\alpha + \frac{1}{2} (f_{yy} - f_{xx}) \sin 2\alpha,$$

de lo cual se sigue inmediatamente que para

$$\cotg 2\alpha = \frac{1}{2} \frac{f_{xx} - f_{yy}}{f_{xy}}$$

tendremos

$$f_{x'y'} = 0.$$

Consideremos ahora la superficie F , dada por la ecuación $z = f(x, y)$, en la cual el origen está en el punto M en consideración, y los ejes Ox y Oy están elegidos en el plano tangente tal que

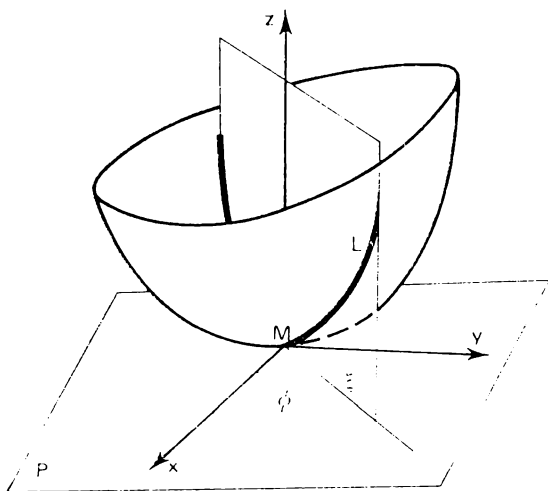


FIG. 27

$f_{xy}(0, 0) = 0$. Tomemos en el plano P una recta arbitraria que forme un ángulo ϕ con el eje Ox y consideremos la sección normal L en la dirección de esta recta (Fig. 27). De la fórmula obtenida en §2, la curvatura de L en el punto M , teniendo en cuenta su signo, es igual a

$$k_L = \lim_{\xi \rightarrow 0} \frac{2f(x, y)}{\xi^2}.$$

Aquí $f(x, y)$ es la distancia (de nuevo teniendo en cuenta su signo) de un punto de L a la recta elegida. Desarrollando $f(x, y)$ por la fórmula de Taylor (capítulo 2, Sec. 9) y observando que $f_x(0, 0) = f_y(0, 0) = 0$ (ya que los ejes Ox y Oy están en el plano tangente), obtenemos

$$f(x, y) = \frac{1}{2} (f_{xx}x^2 + f_{yy}y^2) + \epsilon(x^2 + y^2),$$

donde $\epsilon \rightarrow 0$ cuando $x \rightarrow 0$, $y \rightarrow 0$. Para un punto de L tenemos $x = \xi \cos \phi$, $y = \xi \sin \phi$, $\xi^2 = x^2 + y^2$ (Fig. 27), y por tanto,

$$k_L = \lim_{\xi \rightarrow 0} \frac{f_{xx}\xi^2 \cos^2 \phi + f_{yy}\xi^2 \sin^2 \phi + 2\epsilon\xi^2}{\xi^2} = f_{xx} \cos^2 \phi + f_{yy} \sin^2 \phi.$$

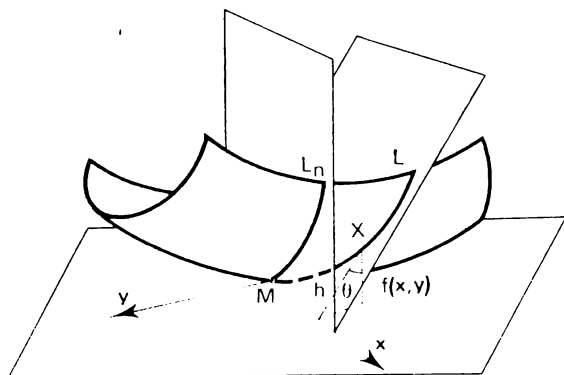


FIG. 28

Poniendo $\phi = 0$, $\phi = \pi/2$, hallamos que f_{xx} y f_{yy} son las curvaturas k_1 y k_2 de las secciones normales en la dirección de los ejes Ox y Oy . De este modo, la fórmula obtenida es, en realidad, la fórmula de Euler: $k = k_1 \cos^2 \phi + k_2 \sin^2 \phi$. El hecho de que k_1 y k_2 sean las curvaturas máxima y mínima también se deduce de esta fórmula.

2. Para la demostración del teorema de Meusnier consideramos una sección normal L_N y una sección L cuyo plano forma un ángulo θ con el plano de L_N , como en la figura 28. Los ejes Ox y Oy están en el plano tangente. Tomamos además el eje Ox tangente a las curvas L_N y L en el origen. La distancia $h(x, y)$ entre el eje Ox y un punto X de L de coordenadas $x, y, f(x, y)$ es evidentemente igual a $h(x, y) = |f(x, y)| / \cos \theta$ (Fig. 28). Utilizando la fórmula de

Taylor, expresamos la curvatura k_L de la curva L de la siguiente manera:

$$\begin{aligned} k_L &= \lim_{x \rightarrow 0} \frac{2h(x, y)}{x^2} = \lim_{x \rightarrow 0} 2 \frac{|f(x, y)|}{x^2 \cos \theta} = \\ &= \lim_{x \rightarrow 0} \frac{|f_{xx}x^2 + 2f_{xy}xy + f_{yy}y^2 + 2\epsilon(x^2 + y^2)|}{x^2 \cos \theta}, \end{aligned} \quad (5)$$

donde $\epsilon \rightarrow 0$ cuando $x, y \rightarrow 0$. Puesto que el eje Ox es tangente a la curva L , evidentemente $\lim_{x \rightarrow 0} y/x = 0$. De este modo, pasando al límite en la fórmula (5), obtenemos

$$k_L = \frac{|f_{xx}|}{\cos \theta}.$$

Pero para el sistema de coordenadas elegido la curva L_N tiene la ecuación $z = f(x, 0)$, por lo que $|k_N| = |f_{xx}|$. Así, pues, $k_L = |k_N|/\cos \theta$, y el teorema de Maunier está probado.

Curvatura media

En muchas cuestiones de la teoría de superficies, el papel más importante está representado no por las curvaturas principales, sino por ciertas cantidades dependientes de ellas, las llamadas *curvatura media* y *curvatura total* o de *Gauss* de la superficie en un punto dado. Examinémoslas con detalle.

La curvatura media de una superficie en un punto es la media aritmética de las curvaturas principales

$$K_m = \frac{1}{2}(k_1 + k_2).$$

Como un ejemplo de la utilidad de este concepto consideremos el siguiente problema de mecánica. Supongamos que sobre la superficie de un cuerpo F hay estirada una membrana de goma elástica tensa. Nos preguntamos cuál es la presión ejercida por esta membrana sobre cada punto de la superficie de F .

La presión en un punto M está medida por la fuerza que ejerce

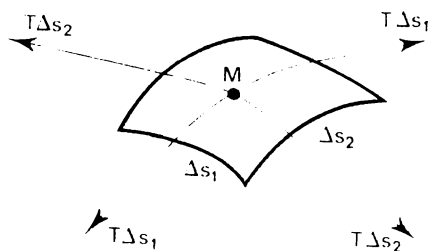


FIG. 29

la membrana sobre un segmento de superficie de área unidad que contenga el punto M ; para ser más exactos, la presión «en el punto» M se define como el límite de la razón de esta fuerza al área del segmento, cuando éste se contrae al punto M .

Rodeemos el punto M de la superficie con un pequeño rectángulo curvilíneo cuyos lados tengan las longitudes Δs_1 e Δs_2 y sean perpendiculares a la primera y segunda direcciones principales en M (figura 29)¹⁶. Sobre cada lado del rectángulo se ejerce una fuerza que es proporcional (por la supuesta uniformidad de la tensión) a la longitud del mismo y a la tensión T que actúa sobre la membrana. De este modo, sobre los lados perpendiculares a la primera dirección principal se ejercen fuerzas que

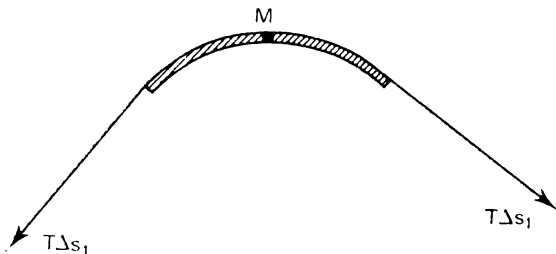


FIG. 30

son aproximadamente iguales a $T\Delta s_1$ y tienen la dirección de la tangente a la superficie. Análogamente, fuerzas iguales a $T\Delta s_2$ actúan sobre el otro par de lados del rectángulo. Para hallar la presión en el punto M , debemos dividir la resultante de estas cuatro fuerzas por el área del rectángulo (aproximadamente igual a $\Delta s_1 \Delta s_2$) y pasar al límite para $\Delta s_1, \Delta s_2 \rightarrow 0$. Empecemos por dividir la resultante de las dos primeras fuerzas por $\Delta s_1 \Delta s_2$.

Si examinamos el rectángulo de canto (Fig. 30), vemos que estas fuerzas están dirigidas según tangentes a la curva de la primera sección normal y que la distancia entre sus puntos de aplicación es exactamente Δs_2 . Se trata, pues, del mismo problema que en §2 para la presión de una cuerda sobre un soporte. Utilizando el anterior resultado, obtenemos que el límite deseado es igual a $k_1 T$, donde k_1 es la curvatura de la primera sección normal. Con una expresión similar para las otras dos fuerzas, obtenemos la fórmula:

$$P_M = T(k_1 + k_2) = 2TK_m.$$

¹⁶ El presente razonamiento no es riguroso. Sin embargo, haciendo estimaciones de los errores introducidos, es posible dar una demostración rigurosa del resultado.

Este resultado tiene muchas consecuencias importantes. Consideremos un ejemplo.

Sabemos que la película superficial de un líquido está bajo una tensión que es la misma en todas direcciones sobre la superficie. Para una masa de líquido limitada por una superficie curva, esta tensión, por el resultado anterior, ejerce sobre la superficie una presión que es proporcional a su curvatura media en el punto en cuestión.

Así, en gotas de diámetro muy pequeño las presiones son muy grandes, hecho que dificulta la formación de tales gotas. Al enfriarse un vapor, las gotas comienzan a formarse normalmente alrededor de motitas de polvo y de partículas cargadas. Enfriando suavemente vapor completamente puro, la formación de gotas se retrasa. Pero si, por ejemplo, una partícula atraviesa el vapor con gran rapidez, causando la ionización de las moléculas, entonces alrededor de los iones formados en su camino aparecerán momentáneamente pequeñas gotas de vapor, constituyendo un rastro visible de la partícula. Esta es la base para la construcción de la cámara de Wilson, muy usada en física nuclear para observar los movimientos de diversas partículas cargadas.

Como la presión ejercida por un líquido es la misma en todas direcciones, una gota de líquido en ausencia de otras fuentes de presión debe adoptar una forma para la cual en todos los puntos de la superficie la curvatura media sea la misma. En el experimento de Plateau se toman dos líquidos del mismo peso específico, con lo cual un grumo de uno de ellos flotará en equilibrio en el otro. Se puede suponer que sobre el líquido flotante solamente actúa la tensión superficial¹⁷, y esto hace que el líquido «flotante» siempre tome la forma de una esfera. Este resultado sugiere que toda superficie cerrada con curvatura media constante es una esfera, teorema que en efecto es cierto, aunque la demostración matemática rigurosa es muy difícil.

También cabe abordar la cuestión desde otro lado. Dado que la tensión superficial tiende a disminuir el área de la superficie, mientras que el volumen de líquido no puede cambiar, es natural esperar que la masa flotante de líquido tenga la menor superficie posible para un volumen dado. Se puede probar que un cuerpo con esta propiedad es una esfera.

¹⁷ El aumento de presión con la profundidad se puede despreciar, ya que es el mismo para ambos líquidos, debido a que tienen el mismo peso específico. Así, sobre su frontera común, las presiones adicionales, interna y externa, causadas por la profundidad, se neutralizan entre sí.

La relación entre la presión lateral de la película y su curvatura media puede también ser utilizada para determinar la forma de una película de jabón suspendida en un perfil. Como a la presión lateral, dirigida según la normal a la superficie, no se le opone ninguna reacción del soporte (el soporte en este caso simplemente no existe), dicha presión debe ser igual a cero, así que para la superficie deseada tenemos la condición

$$K_m = 0. \quad (6)$$

De la expresión analítica para la curvatura media obtenemos una ecuación diferencial, y el problema consiste en resolver esta ecuación con la condición de que la superficie buscada pase a través del perfil dado¹⁸. Han sido muchas las investigaciones sobre este difícil problema.

La misma ecuación (6) aparece en el problema de hallar la superficie de menor área limitada por un contorno dado. Desde un punto de vista físico, la identidad de estos dos problemas es natural, ya que la película tiende a disminuir su área y alcanza una posición de equilibrio estable cuando logra el área mínima posible bajo las condiciones dadas. Debido a su conexión con este problema, las superficies de curvatura media nula se denominan *mínimas*.

La investigación matemática de superficies mínimas es de gran interés, en parte debido a la gran variedad de formas esencialmente diferentes, como se ha descubierto a través de experimentos con películas de jabón. La figura 31 muestra dos películas de jabón suspendidas en diferentes contornos.

Curvatura de Gauss

La *curvatura de Gauss* de una superficie en un punto dado es el producto de las curvaturas principales

$$K = k_1 k_2.$$

El signo de la curvatura de Gauss determina el carácter de la superficie en las proximidades del punto considerado. Para $K > 0$ la superficie tiene la forma de un cuenco (k_1 y k_2 tienen el mismo signo) y para $K < 0$, cuando k_1 y k_2 tienen signos diferentes, la super-

¹⁸ Para una superficie dada por la ecuación $z = z(x, y)$, la ecuación (6) toma la forma

$$(1 + z_y'^2)z_{xx}'' - 2z_x'z_y'z_{xy}'' + (1 + z_x'^2)z_{yy}'' = 0.$$

ficie es como una silla de montar. Los restantes casos, discutidos anteriormente, corresponden a la curvatura de Gauss nula. El valor absoluto de la curvatura de Gauss da el grado de curvatura de la superficie en general, como un tipo de abstracción de las distintas curvaturas en diferentes direcciones. Esto se hace particularmente claro si consideramos una definición diferente de la curvatura de Gauss que no dependa del estudio de curvas sobre la superficie:

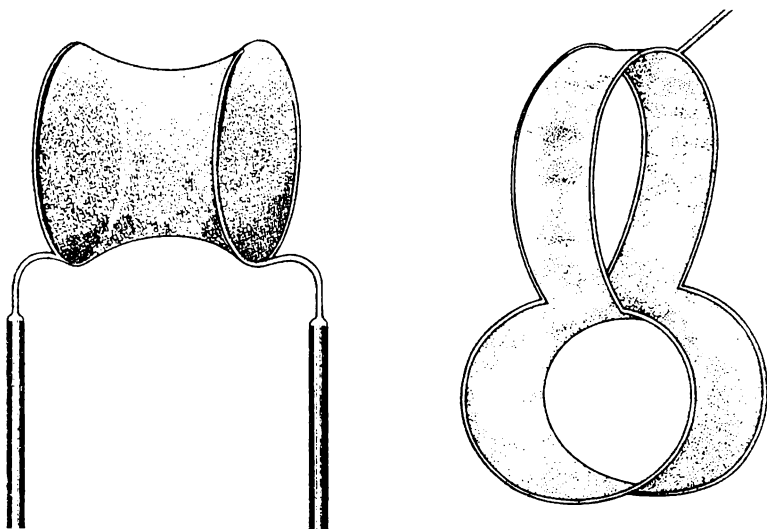


FIG. 31

Consideremos un pequeño segmento G de la superficie F , que contenga el punto M en su interior, y en cada punto de este segmento tracemos una normal a la superficie.

Si trasladamos los puntos iniciales de todas estas normales a un punto, entonces estas normales llenan un ángulo sólido (Fig. 32). El valor de este ángulo sólido dependerá del área del segmento G y del grado en que la superficie esté curvada en este segmento. Así, pues, el grado de curvatura del segmento G puede ser caracterizado por la razón del valor del ángulo sólido al área de G ; por tanto, es natural definir la curvatura de la superficie en un punto dado como

el límite de esta razón cuando el segmento G se contrae al punto M ¹⁹. Se obtiene que este límite es igual al valor absoluto de la curvatura de Gauss en el punto M .

La propiedad más notable de la curvatura de Gauss (propiedad que explica su gran significado en la teoría de superficies) es la siguiente. Supongamos que la superficie es de un material flexible

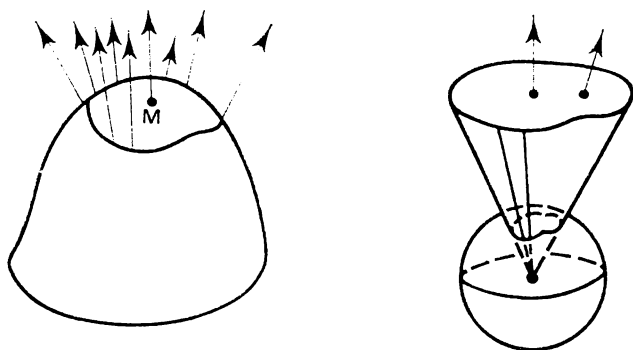


FIG. 32

pero inextensible —por ejemplo, una hoja muy delgada de estaño—, de modo que podemos combarla de distintas formas sin estirla ni romperla. Durante este proceso las curvaturas principales cambiarán, pero, como probó Gauss, el producto $k_1 k_2$ permanecerá invariable en cada punto. Este resultado fundamental prueba que dos superficies con diferentes curvaturas de Gauss son inherentemente distintas entre sí, consistiendo esta diferencia en el hecho de que si las deformamos de todas las formas posibles, sin estirlas ni romperlas, nunca podremos superponer una sobre la otra. Por ejemplo, un segmento de la superficie de una esfera nunca se podrá deformar de tal forma que se adapte a un plano o a la superficie de una esfera de distinto radio.

Hasta aquí hemos considerado ciertos conceptos básicos en la teoría de superficies. En cuanto a los métodos usados en esta teoría

¹⁹ Para medir el ángulo sólido, construimos una esfera de radio unidad con centro en su vértice. El área de la región en la que la esfera corta al ángulo sólido se toma entonces como la dimensión del ángulo sólido (Fig. 32).

consisten, como ya se dijo con anterioridad, fundamentalmente en la aplicación del análisis y sobre todo de las ecuaciones diferenciales. Ejemplos sencillos del uso del análisis son las demostraciones de los teoremas de Euler y Meusnier. Para cuestiones más complicadas, necesitamos un método especial de relacionar los problemas de la teoría de superficies con los problemas del análisis. Este método está basado en la introducción de las llamadas coordenadas curvilíneas y fue por primera vez utilizado ampliamente en el trabajo de Gauss sobre problemas del tipo discutido en la sección siguiente.

§4. *Geometría intrínseca y deformación de superficies*

Geometría intrínseca

Como se indicó anteriormente, una deformación de una superficie se define como un cambio de forma que conserva las longitudes de todas las curvas sobre la superficie. Por ejemplo, enrollar una hoja de papel sobre un tubo cilíndrico representa, desde el punto de vista geométrico, una deformación de parte del plano, ya que el papel no experimenta prácticamente alargamiento alguno, y la longitud de cualquier curva dibujada sobre él no cambia en el proceso. Algunas otras cantidades geométricas relacionadas con la superficie también se conservan; por ejemplo, el área de figuras sobre ella. Todas las propiedades de una superficie que no cambian en el curso de una deformación constituyen la llamada *geometría intrínseca* de la superficie.

Pero ¿cuáles son estas propiedades? Está claro que en una deformación solamente se pueden conservar aquellas propiedades que en definitiva dependan por entero de la longitud de curvas, es decir, que se puedan determinar a través de medidas hechas sobre la propia superficie. Una deformación es un cambio de forma que conserva la longitud de las curvas, y cualquier propiedad que no pueda cambiar bajo *ninguna* deformación debe ser definible de una forma u otra en términos de longitud. De este modo, la geometría intrínseca se denomina simplemente *geometría sobre una superficie*. El verdadero sentido de las palabras «geometría intrínseca» es que estudia propiedades intrínsecas de la superficie, independientemente de la manera como ésta esté sumergida en el espacio circundante ²⁰. Así, por ejem-

²⁰ Observemos que las ideas de la geometría intrínseca han conducido a una amplia generalización del concepto matemático de espacio y han representado, como consecuencia de ello, un importante papel en la física contemporánea; para los detalles véase el capítulo 16.

plo, si unimos dos puntos de una hoja de papel mediante una recta y curvamos luego el papel (Fig. 33), el segmento se convierte en una curva, pero su propiedad de ser la línea más corta de todas las que unen los puntos dados sobre la superficie se conserva; así, pues, esta propiedad pertenece a la geometría intrínseca. Por otra parte, la curvatura de esta línea dependerá de cómo se haya doblado el papel, y por tanto no forma parte de la geometría intrínseca.

En general, puesto que las demostraciones de la geometría plana

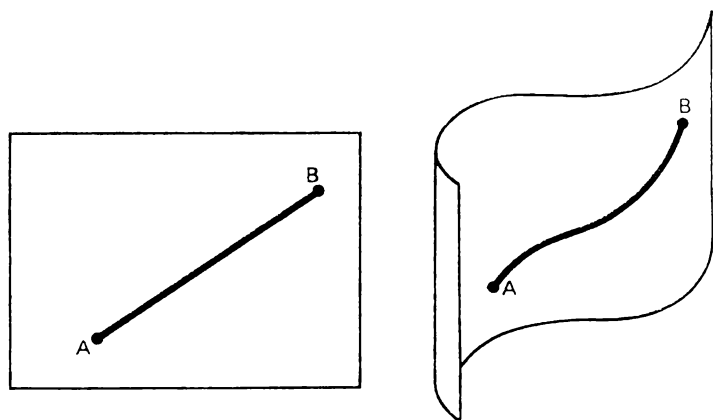


FIG. 33

no hacen referencia al espacio circundante, todos sus teoremas pertenecen a la geometría intrínseca de cualquier superficie obtenible por deformación, a partir de un plano. Se puede decir que la geometría plana es la geometría intrínseca del plano.

Otro ejemplo de geometría intrínseca es conocido de todos: la geometría sobre la superficie de una esfera, que es la que se utiliza normalmente al hacer medidas sobre la superficie de la Tierra. Este ejemplo es particularmente idóneo para ilustrar la naturaleza esencial de la geometría intrínseca; debido al gran radio de la Tierra, cualquier área directamente abaricable con la vista nos da la impresión de ser parte de un plano, así que las desviaciones a partir de la geometría plana que se observan en las medidas de grandes distancias nos dan la impresión de que resultan no de la curvatura de la superficie de la tierra en el espacio, sino de las leyes inherentes a la «geo-

metría terrestre», que expresa las propiedades geométricas de la Tierra misma.

Queda por señalar que la idea de estudiar la geometría intrínseca se le ocurrió a Gauss en relación con los problemas de geodesia y cartografía. Ambas ciencias aplicadas se refieren de manera esencial a la geometría intrínseca de la superficie de la tierra. La cartografía trata, en particular, sobre distorsiones en la razón de distancias cuando se proyecta parte de la superficie de la tierra sobre un plano;

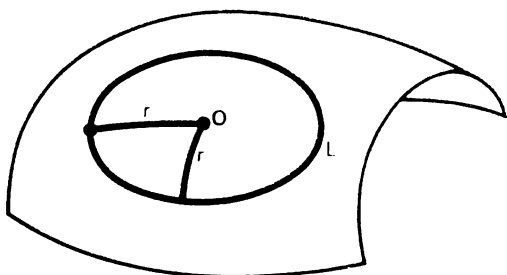


FIG. 34

trata, pues, sobre la distinción entre geometría plana y geometría intrínseca de la superficie de la tierra.

La geometría intrínseca de cualquier otra superficie se puede describir de la misma manera. Imaginemos que sobre una superficie dada existen seres tan pequeños que dentro de los límites visuales la superficie parezca un plano (sabemos que un segmento suficientemente pequeño de una superficie regular se diferencia muy poco de un plano tangente); entonces estos seres no advertirán que la superficie está curvada en el espacio, pero midiendo largas distancias se convencerán, no obstante, de que en su geometría prevalecen leyes que no corresponden a un plano, y sí, en cambio, a la geometría intrínseca de la superficie en que ellos viven. Que estas leyes son realmente diferentes para distintas superficies se puede ver con facilidad en la siguiente discusión. Tomemos un punto O sobre la superficie y consideremos una curva L tal que la distancia de cada uno de sus puntos a O , medida sobre la superficie (es decir, a lo largo de la curva más corta que una el punto con O) sea igual a un número fijo r (Fig. 34). La curva L , desde el punto de vista de la geometría intrínseca de la superficie, es simplemente la circunferencia de un círculo

de radio r . Una fórmula que exprese la longitud $s(r)$ en función de r forma parte de la geometría intrínseca de la superficie dada. Sin embargo, el carácter de tal fórmula puede variar grandemente según sea la naturaleza de la superficie: así, sobre un plano, $s(r) = 2\pi r$; sobre una esfera de radio R , como puede probarse fácilmente, $s(r) = 2\pi R \sin r/R$; sobre la superficie de la figura 35, y empezando con un cierto valor de r , la longitud de la circunferencia de centro O y radio r es, al principio, independiente de r , para después comenzar a decrecer. En consecuencia, todas estas superficies tienen diferentes geometrías intrínsecas.

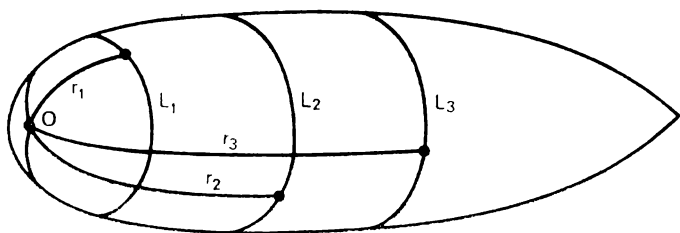


FIG. 35

Los conceptos básicos de la geometría intrínseca

Para ilustrar la amplia gama de conceptos y teoremas de la geometría intrínseca podemos volver a la geometría plana que, como hemos visto, es la geometría intrínseca del plano. La geometría plana trata de figuras del plano y de sus propiedades; éstas vienen expresadas normalmente en forma de relaciones entre magnitudes geométricas básicas tales como longitud, ángulo y área. Para una demostración rigurosa de que los ángulos y las áreas pertenecen a la geometría intrínseca del plano es necesario probar que se pueden expresar en términos de longitud. Esto es efectivamente así, pues un ángulo se puede calcular si se conoce la longitud de los lados de un triángulo que lo contenga, y el área de un triángulo se puede también calcular en función de sus lados, mientras que para hallar el área de un polígono solamente necesitamos descomponerlo en triángulos.

Al considerar la geometría plana como la geometría intrínseca del plano, no necesitamos restringirnos a las ideas aprendidas en la escuela. Por el contrario, podemos desarrollar dicha geometría tanto como gustemos y estudiar numerosos problemas nuevos, con la única condición de que puedan ser enunciados, en definitiva, en términos de

longitud. De este modo, en geometría plana podemos introducir sucesivamente la longitud de una curva, el área de una superficie limitada por curvas, etc.; todo esto forma parte de la geometría intrínseca del plano.

Los mismos conceptos se introducen en la geometría intrínseca de una superficie arbitraria. La longitud de una curva es el concepto inicial; la definición de ángulos y áreas es algo más complicada. Si la geometría intrínseca de una superficie dada difiere de la plana, no podemos utilizar las fórmulas usuales para definir un ángulo o

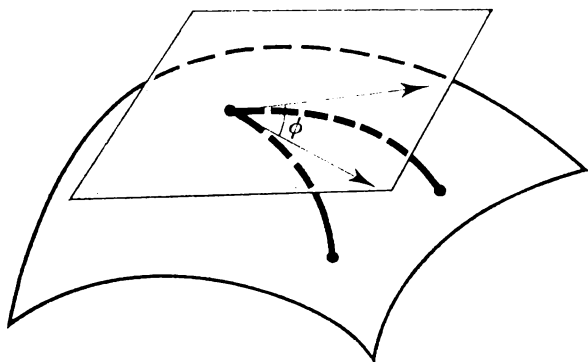


FIG. 36

un área en términos de longitud. Sin embargo, como hemos visto, una superficie en las proximidades de un punto dado se diferencia poco de su plano tangente. Hablando con más precisión, podemos decir: si se proyecta un pequeño segmento de una superficie que contenga un punto dado M sobre el plano tangente en este punto, entonces la distancia entre dos puntos, medida sobre la superficie, difiere de la distancia entre sus proyecciones en un infinitésimo de orden mayor que dos en comparación con las distancias al punto M . Así, pues, al definir cantidades geométricas en un punto dado de una superficie tomando un límite en el que aparezcan infinitésimos de orden no superior al segundo, es lícito reemplazar un segmento de la superficie por su proyección sobre el plano tangente. De este modo, las cantidades medidas en el plano tangente resultan formar parte de la geometría intrínseca de la superficie. Esta posibilidad de considerar un pequeño segmento de la superficie como un plano es la base de las definiciones de todos los conceptos de la geometría intrínseca.

Como ejemplo, consideremos las definiciones de ángulo y área. Siguiendo el principio general, definimos el ángulo entre curvas sobre una superficie como el ángulo entre sus proyecciones sobre el plano tangente (Fig. 36). Evidentemente, el ángulo definido de esta forma es idéntico al formado por las tangentes a las curvas. La definición de área dada en §3 está basada en el mismo principio. Finalmente, para que la tendencia de una curva a torcerse en el espacio pueda definirse «dentro» de la misma superficie, introducimos el concepto de «curva geodésica», nombre que recuerda a las medidas efectuadas sobre la superficie de la Tierra. La *curvatura geodésica* de una curva en un punto se define como la curvatura de su proyección sobre el plano tangente (Fig. 37).

De esta forma vemos que los conceptos básicos de la geometría plana se pueden transferir a la geometría intrínseca de una superficie arbitraria.

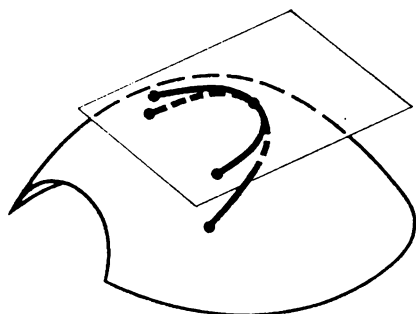


FIG. 37

En una superficie arbitraria es también fácil definir figuras análogas a las figuras básicas sobre el plano. Por ejemplo, antes hablamos ya de circunferencias de círculos, que se definieron precisamente como en el caso del plano. Análogamente podemos definir el equivalente de un segmento de recta, un segmento geodésico, como la curva más corta (sobre la superficie) que une dos puntos. Por otro lado, es natural definir un triángulo como la figura limitada por tres segmentos geodésicos y lo mismo para un polígono, etc. Como las propiedades de todas estas figuras y magnitudes dependen de la superficie, existen en este sentido infinitas geometrías intrínsecas diferentes. Mas la geometría intrínseca, como rama especial de la teoría de superficies, presta particular atención a ciertas leyes generales pertenecientes a la geometría intrínseca de cualquier superficie y pone en claro la manera como se expresan dichas leyes en términos de las magnitudes que caracterizan a una superficie dada.

Así, pues, como ya hemos observado anteriormente, una de las más importantes características de una superficie, la curvatura de Gauss, no varía por deformación, es decir, depende solamente de la

geometría intrínseca de la superficie. Pero ocurre que, en general, la curvatura de Gauss caracteriza ya, en alto grado, la medida en que la geometría intrínseca de la superficie difiere, en las proximidades de un punto dado, de la geometría plana. Como ejemplo consideremos sobre una superficie un círculo L de radio muy pequeño r , con centro en un punto O . Sobre un plano la longitud $s(r)$ de su circunferencia viene expresada por la fórmula $s(r) = 2\pi r$. Sobre una superficie diferente de un plano, la dependencia de la circunferencia con respecto al radio es distinta; aquí la desviación de $s(r)$ a partir de $2\pi r$ depende esencialmente, para r pequeño, de la curvatura de Gauss K en el centro del círculo, es decir:

$$s(r) = 2\pi r - \frac{\pi}{3} Kr^3 + \epsilon r^3,$$

donde $\epsilon \rightarrow 0$ cuando $r \rightarrow 0$. Dicho con otras palabras: para r pequeño, la circunferencia se puede calcular por la fórmula usual si despreciamos los términos de tercer grado de pequeñez, y en este caso el error

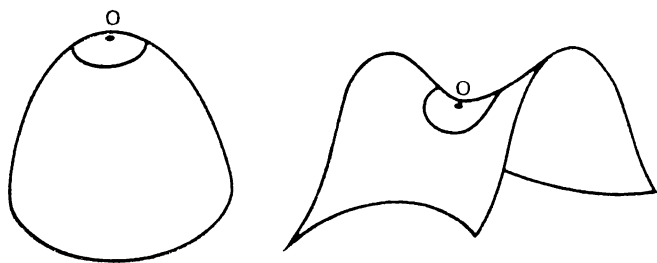


FIG. 38

(con una aproximación superior al tercer orden) es proporcional a la curvatura de Gauss. En particular, si $K > 0$, entonces la circunferencia de un círculo de radio pequeño es menor que la circunferencia de un círculo con el mismo radio en el plano; y si $K < 0$, es mayor. Esto último se ve con facilidad: en las proximidades de un punto con curvatura positiva la superficie tiene la forma de un cuenco, por lo que las circunferencias se reducen, mientras que en las proximidades de un punto con curvatura negativa la circunferencia, situada sobre una «silla de montar», tiene forma ondulada, y por tanto es considerablemente más larga (Fig. 38).

Del teorema que acabamos de mencionar se deduce que una superficie con curvatura de Gauss variable es extremadamente heterogénea desde un punto de vista geométrico; las propiedades de su geometría intrínseca cambian de un punto a otro. El carácter general de los problemas de geometría intrínseca hace que ésta se parezca a la geometría plana; pero dicha heterogeneidad, por otra parte, la hace profundamente diferente a ésta. Sobre el plano, por ejemplo, la suma de los ángulos de un triángulo es igual a dos rectos; en cambio, sobre una superficie arbitraria, la suma de los ángulos de un triángulo (con geodésicas por lados) es indeterminada, incluso sabiendo sobre qué superficie se halla el triángulo y conociendo los lados del mismo. Sin embargo, si conocemos la curvatura de Gauss K en cada punto del triángulo, entonces la suma de sus ángulos, α, β, γ se puede calcular mediante la fórmula

$$\alpha + \beta + \gamma = \pi + \iint K d\sigma,$$

donde la integral está extendida a la superficie del triángulo. Esta fórmula contiene como caso especial los conocidos teoremas relativos a la suma de los ángulos de un triángulo en el plano y sobre la esfera unidad. En el primer caso $K = 0$, y $\alpha + \beta + \gamma = \pi$; mientras que en el segundo $K = 1$, y $\alpha + \beta + \gamma = \pi + S$, donde S es el área del triángulo esférico.

Se puede probar que todo segmento suficientemente pequeño de una superficie con curvatura de Gauss nula se puede deformar —o, como se acostumbra a decir, desarrollar en un plano—, ya que tiene la misma geometría intrínseca que éste. Tales superficies se denominan *desarrollables*. Y si la curvatura de Gauss es próxima a cero, entonces, aunque la superficie no se pueda desarrollar en un plano, no obstante su geometría intrínseca difiere muy poco de la geometría plana, lo cual indica una vez más que la curvatura de Gauss viene a ser una medida del grado en que la geometría intrínseca de una superficie se desvía de la plana.

Líneas geodésicas

En la geometría intrínseca de una superficie, el papel de las líneas rectas está representado por las líneas geodésicas, o, como se denominan normalmente, «geodésicas».

Una recta en el plano se puede definir como una línea formada por intervalos que se solapan entre sí. Una geodésica se define exactamente de la misma forma, con segmentos geodésicos en lugar de

intervalos. En otras palabras, una *geodésica* es una curva sobre una superficie, tal que cada trozo suficientemente pequeño de ella es el camino más corto entre dos puntos. No toda geodésica es, globalmente, un camino mínimo, como puede observarse sobre la superficie de una esfera, donde todo arco de un círculo máximo es una geodésica, aunque este arco sólo será el camino más corto entre sus extremos si no es mayor que un semicírculo. Una geodésica, como vemos; puede incluso ser una curva cerrada.

Para ilustrar ciertas propiedades importantes de las geodésicas, consideremos el siguiente modelo mecánico ²¹.

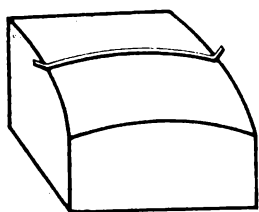


FIG. 39

Imaginemos sobre la superficie F una cuerda de goma tirante, con los extremos fijos (Fig. 39) ²². La cuerda estará en equilibrio cuando tenga la longitud más corta posible, ya que cualquier cambio en su posición llevaría consigo un aumento de longitud, que sólo podría ser producido por fuerzas externas. Dicho con otras palabras, la cuerda estará en equilibrio si está tendida a lo largo de una geodésica. Mas para el equilibrio es necesario que las fuerzas elásticas sobre cada

segmento estén compensadas por la resistencia de la superficie, dirigida según la normal. (Suponemos que la superficie es regular y que no hay rozamiento entre ella y la cuerda.) Ahora bien, probamos en §2 que la presión sobre el soporte causada por la tensión de la cuerda está dirigida según la normal principal a la curva sobre la cual yace la cuerda. Llegamos así al siguiente resultado: la normal principal a una geodésica en cada punto coincide en dirección con la normal a la superficie. El recíproco de este teorema es también cierto: toda curva sobre una superficie regular que tenga esta propiedad es una geodésica.

Esta propiedad de una geodésica nos permite deducir el siguiente hecho importante: si un punto material se mueve sobre una superficie de tal forma que sobre él no actúen más fuerzas que la de la reacción de la superficie entonces dicho punto sigue una geodésica. Puesto que,

²¹ Como hemos observado previamente, nuestro razonamiento no es una demostración estricta de las propiedades de las curvas geodésicas. Se da sólo para ilustrar las más importantes de estas propiedades.

²² En general, una cuerda estirada no se apoya toda ella sobre una superficie a menos que ésta sea convexa; así, para no hacer excepciones, es mejor imaginar que la superficie tiene dos caras y que la cuerda pasa entre ambas.

como sabemos de §2, la aceleración normal de un punto está dirigida según la normal principal a la trayectoria, y como la reacción de la superficie es la única fuerza que actúa sobre el punto, la normal principal a la trayectoria coincide con la normal a la superficie, de modo que, por el teorema precedente, la trayectoria es una geodésica. Esta última propiedad de las geodésicas aumenta su semejanza con las rectas. Igual que el movimiento de un punto libre, como consecuencia de la inercia, se produce a lo largo de una recta, así el movimiento de un punto obligado a apoyarse sobre una superficie, pero no afectado por fuerzas externas, tendrá lugar a lo largo de una geodésica ²³.

De la misma propiedad de las geodésicas proviene el siguiente teorema. Si dos superficies son tangentes a lo largo de una curva que es una geodésica sobre una de ellas, entonces esta curva también será una geodésica sobre la otra. Puesto que en cada punto de la curva las superficies tienen un plano tangente común y, por consiguiente, una normal común, y como la curva es una geodésica sobre una de las superficies, esta normal coincide con la normal principal a la curva, de modo que sobre la segunda superficie esta curva también será una geodésica.

De estos resultados se siguen dos nuevas propiedades intuitivas de las curvas geodésicas. En primer lugar, si una placa rectangular elástica (por ejemplo, una regla de acero) se apoya con su línea media completamente sobre una superficie, entonces es tangente a ésta a lo largo de una geodésica. (Evidentemente, la línea de contacto es una geodésica sobre la regla, así que debe serlo también sobre la superficie.) En segundo lugar, si una superficie rueda sobre un plano de tal forma que el punto de contacto traza una recta sobre el plano entonces la traza de esta recta sobre la superficie es una geodésica ²⁴. Ambas propiedades se demuestran fácilmente sobre un cilindro. En relación con esta figura es fácil convencerse mediante experimentos de que la línea media de una cinta plana recta apoyada sobre el cilindro (Fig. 40) coincide o con una generatriz, o con la circunferencia de un círculo, o con una hélice, y no es difícil probar que una línea geodésica sobre un cilindro puede ser solamente una de estas tres. Las mismas curvas se marcarán sobre un cilindro si lo hacemos rodar sobre un plano en el que hemos dibujado una recta con tiza.

²³ Por fuerzas «externas» entendemos aquí todas las fuerzas salvo la reacción de la superficie.

²⁴ Esta proposición no difiere esencialmente de la precedente, ya que el rodamiento de una superficie sobre un plano es equivalente, en un sentido bien determinado, a desenrollar una banda plana a lo largo de la superficie.

La analogía entre geodésicas y rectas en un plano puede ser completada con otra importante propiedad, tomada directamente de la definición de geodésica: a saber, que las rectas del plano se pueden definir como curvas de curvatura cero y las geodésicas de una superficie como curvas de curvatura geodésica cero. (Recordemos que la curvatura geodésica es la curvatura de la proyección de la curva sobre el plano tangente, figura 37.) Es completamente natural que nuestra presente definición de geodésica coincida con la anterior, puesto que,

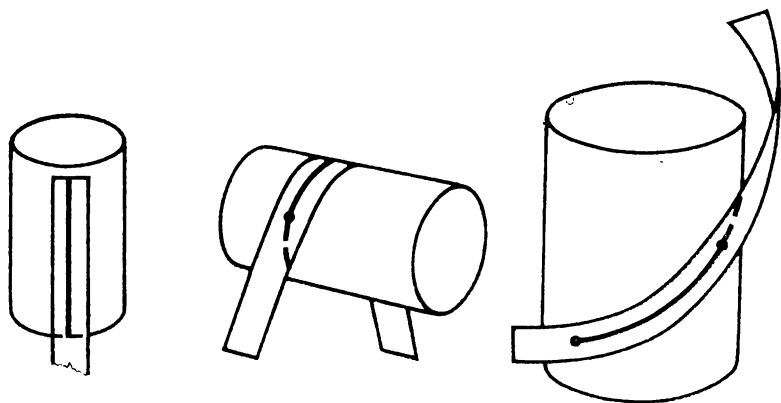


FIG. 40

si en todo punto de la curva la curvatura de su proyección sobre el plano tangente es igual a cero, entonces la curva parte de su tangente esencialmente en la dirección de la normal a la superficie, de modo que la normal principal a la curva está dirigida según la normal a la superficie y la curva es una geodésica en el sentido original. Recíprocamente, si una curva es una geodésica, entonces su normal principal —y por tanto también su desviación de la tangente— está dirigida según la normal a la superficie, de modo que, proyectándola sobre el plano tangente, obtenemos una curva en la cual la desviación a partir de la tangente es esencialmente menor que para la curva original, y la curvatura de la proyección así formada es igual a cero.

El camino de una geodésica puede variar mucho para distintas superficies. En la figura 41 aparecen, a título de ejemplo, algunas geodésicas sobre un hiperboloide de revolución.

Deformación de superficies

Toda vez que la geometría intrínseca estudia las propiedades de superficies que son invariantes bajo una deformación, es natural que investigue también estas deformaciones. La teoría de la deformación de superficies es una de las más interesantes y difíciles ramas de la geometría, e incluye muchos problemas que, aunque sencillos de enunciar, no han sido todavía resueltos definitivamente.

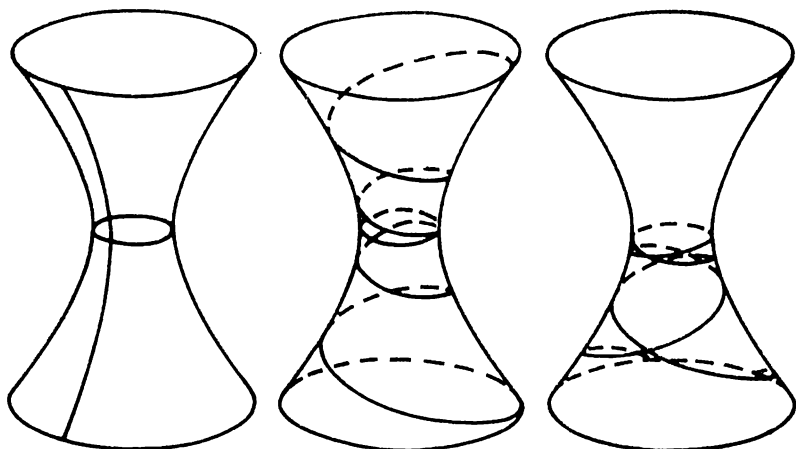


FIG. 41

Ciertas cuestiones sobre la deformación de superficies fueron ya consideradas por Euler y Minding, pero resultados generales para superficies arbitrarias no se obtuvieron hasta más tarde.

En la teoría general de la deformación planteamos en primer lugar la cuestión de si la deformación es posible para todas las superficies, y si es así, hasta qué punto. Para superficies analíticas, es decir, superficies definidas por funciones de las coordenadas que pueden ser desarrolladas en serie de Taylor, esta cuestión fue resuelta a finales del siglo pasado por el matemático francés Darboux. En particular, probó lo siguiente: si sobre una superficie de este tipo consideramos una geodésica y fijamos en el espacio una curva (analítica) arbitraria de la misma longitud y con curvatura no nula en todos los puntos, entonces una banda suficientemente estrecha de la superficie y que

contenga a la geodésica dada se puede deformar de modo que ésta coincida con la curva dada ²⁵. Este teorema muestra que una banda de una superficie se puede deformar un tanto arbitrariamente. Sin embargo, se ha demostrado que, si lo que se desea es transformar una geodésica en una curva prefijada, entonces la superficie se puede deformar a lo más de dos maneras. Por ejemplo, si la curva es plana, entonces las dos posiciones de la superficie serán imágenes especulares una de otra en el plano. Si la geodésica es una recta, entonces la última proposición no es cierta, como se puede probar deformando una superficie cilíndrica.

Hemos definido el concepto de deformación como una transformación de la superficie que conserva las longitudes de todas las curvas sobre ella. Aquí hemos considerado solamente el resultado final de la transformación; la cuestión de qué ocurre con la curva durante el proceso no ha sido examinada. Sin embargo, al considerar una superficie como hecha de un material flexible pero inextensible, es natural examinar una transformación continua en cada instante de la cual las longitudes permanezcan constantes (físicamente esto corresponde a la inextensibilidad del material). Tales transformaciones se denominan deformaciones *continuas*.

A primera vista puede parecer que toda deformación se puede realizar de una manera continua, pero esto no es así. Por ejemplo, se ha probado que una superficie en forma de un canalón circular (Fig. 42), no admite deformaciones continuas (esto explica, entre

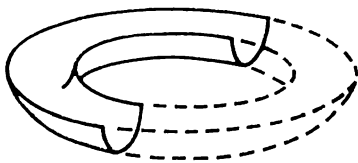


FIG. 42

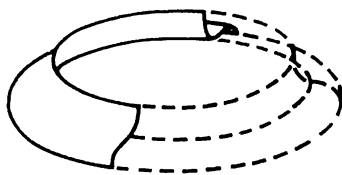


FIG 43

otras cosas, el hecho familiar de que un cubo de bordes curvados es considerablemente más fuerte que otro de borde plano), aunque las deformaciones de tal superficie sí son posibles: por ejemplo, se puede cortar el canalón a lo largo de la circunferencia sobre la que se apoya

²⁵ El caso de transformar una geodésica en una curva de curvatura cero se excluye, ya que es fácil probar que para superficies de curvatura de Gauss positiva esto es imposible.

en un plano horizontal y sustituir una mitad por su imagen especular con respecto a este plano (compárese la figura 43 con la 42; para una mejor visualización hemos dibujado solamente la mitad izquierda de la superficie). Intuitivamente está claro que la imposibilidad de una deformación continua es debida a la forma circular del canalón; para un canalón recto tal deformación puede efectuarse continuamente.

Si nos restringimos a un segmento suficientemente pequeño de la superficie, entonces no existen obstáculos para su deformación continua, y cabría esperar que toda deformación de un pequeño segmento de la superficie pudiera ser realizada mediante una transformación continua, seguida quizá de una simetría con respecto a un plano. Esto es, en efecto, cierto, pero sólo con la condición de que sobre el pequeño segmento de la superficie la curvatura de Gauss nunca se anule (exceptuando el caso en que sea nula en todos los puntos). Por el contrario, si la curvatura de Gauss se anula en puntos aislados, entonces, como demostró N. V. Efimov en 1940, incluso segmentos arbitrariamente pequeños de una superficie regular pueden no admitir una deformación continua sin pérdida de regularidad. Por ejemplo, la superficie definida por la ecuación $z = x^9 + \lambda x^6 y^3 + y^9$, donde λ es un número trascendente, tiene la propiedad de que ningún segmento que contenga el origen, por pequeño que aquél sea, admite deformaciones continuas suficientemente regulares. El teorema de Efimov constituye un resultado nuevo y algo inesperado en la geometría diferencial clásica.

Además de estas cuestiones generales sobre deformación, se está presentando mucha atención a tipos especiales de deformación de superficies.

Conexión entre la geometría intrínseca de una superficie y la forma de ésta en el espacio

Ya sabemos que ciertas propiedades de una superficie, y de las figuras sobre ella, están determinadas por la geometría intrínseca de la misma, aun cuando estas propiedades estén muy estrechamente relacionadas con otras que dependen de cómo está sumergida la superficie en el espacio circundante, propiedades que son, como suele decirse, «extrínsecas» a la superficie. Por ejemplo, las curvaturas principales son propiedades extrínsecas, pero su producto (la curvatura de Gauss) es intrínseca. Otro ejemplo: para que la normal principal a una curva sobre la superficie coincida con la normal a ésta, es necesario y suficiente que esta curva tenga una propiedad que viene determinada por su geometría intrínseca, a saber, que sea una geodésica.

Por consiguiente, la geometría intrínseca de una superficie determina su forma en el espacio sólo hasta cierto punto.

La dependencia de la forma de una superficie en el espacio con respecto a su geometría intrínseca se puede expresar analíticamente en forma de ecuaciones que contienen ciertas cantidades que caracterizan dicha geometría, y otras que caracterizan la forma en que la superficie curvada está sumergida en el espacio. Una de estas ecuaciones, la fórmula que expresa la curvatura de Gauss en términos intrínsecos, es debida a Gauss. Otras dos ecuaciones de este tipo son las de Peterson y Codazzi, mencionadas en la sección 1.

Las ecuaciones de Gauss, Peterson y Codazzi expresan completamente la conexión entre la geometría intrínseca de una superficie y el carácter de su curvatura en el espacio, ya que todas las posibles interrelaciones entre propiedades intrínsecas y extrínsecas de una superficie arbitraria están incluidas, al menos en forma implícita, en estas ecuaciones.

Puesto que la forma de una superficie en el espacio no está totalmente determinada por su geometría intrínseca, naturalmente nos preguntamos: ¿qué propiedades extrínsecas debemos tener en cuenta además para determinar totalmente la superficie? Se comprueba que, si dos superficies tienen la misma geometría intrínseca y si, en puntos y direcciones correspondientes, las curvaturas de las secciones normales de estas superficies tienen el mismo signo, entonces son congruentes; es decir, mediante una traslación se puede conseguir que una coincida con la otra. Observemos que Peterson descubrió este teorema quince años antes que Bonnet, con cuyo nombre, sin embargo, se asocia normalmente.

Aparato analítico en la teoría de superficies

La aplicación sistemática del análisis a la teoría de superficies llevó a la construcción de un aparato analítico especialmente apropiado para este propósito. El paso decisivo en esta dirección fue dado por Gauss, quien introdujo el método de representación de superficies mediante las llamadas coordenadas curvilíneas. Este método es una generalización natural de la idea de las coordenadas cartesianas sobre el plano y está estrechamente conectado con la geometría intrínseca de la superficie, para la cual la representación mediante una ecuación de la forma $z = f(x, y)$ no es conveniente. El inconveniente consiste en que las coordenadas x, y de un punto sobre la superficie varían al deformarse ésta. Para eliminar esta dificultad, se toman las coordenadas sobre la misma superficie; dichas coordenadas determinan cada punto por medio de dos números

u y v , que están asociados al punto dado y permanecen asociados a él incluso después de la deformación de la superficie. Las coordenadas del espacio x, y, z del punto serán en cada caso funciones de u y v . Estos números que determinan el punto sobre una superficie se denominan sus *coordenadas curvilíneas*. La elección del nombre se comprende como sigue: si fijamos el valor de una de estas coordenadas, por ejemplo v , y variamos la otra, entonces obtenemos una curva coordenada sobre la superficie. Las curvas coordenadas forman una

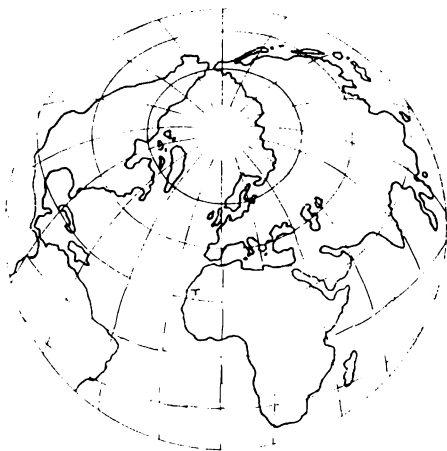


FIG. 44

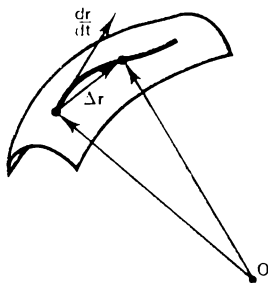


FIG. 45

red curvilínea sobre la superficie similar a la red coordenada sobre el plano. Observemos que el método usual de describir la posición de un punto sobre la superficie de la tierra especificando su longitud y latitud consiste simplemente en la introducción de coordenadas curvilíneas sobre la superficie de una esfera; la red coordenada en este caso está formada por círculos, es decir, los meridianos y paralelos²⁶ (Fig. 44). Para describir la posición espacial de una superficie por medio de coordenadas curvilíneas necesitamos determinar la posición

²⁶ Muy característico es el hecho de que las coordenadas geográficas y sus aplicaciones prácticas se conocieran mucho antes de la introducción por Descartes de las coordenadas usuales en el plano.

de cada punto en términos de u y v , por ejemplo, conociendo en función de u y v el vector $\mathbf{r} = \mathbf{r}(u, v)$, con origen en un punto fijo y extremo en cada punto de la superficie y denominado radio vector de la superficie. (Esto es equivalente a conocer las componentes x , y y z del vector $\mathbf{r}$ como funciones de u y v)²⁷. Para definir una curva sobre una superficie dada, necesitamos conocer las coordenadas u , v como funciones de un parámetro t ; entonces el radio vector de un punto que se mueva a lo largo de esta curva viene expresado como una función compuesta

$$\mathbf{r}[u(t), v(t)].$$

Para funciones vectoriales, los conceptos de derivada y diferencial se pueden generalizar palabra por palabra; de la definición de derivada como el límite de $\Delta \mathbf{r} / \Delta t$ cuando $\Delta t \rightarrow 0$ ($\mathbf{r}$ es una función del parámetro t) se deduce inmediatamente que la derivada del radio vector de una curva es un vector dirigido según la tangente a ella (Fig. 45). Para funciones vectoriales las propiedades básicas de las derivadas ordinarias son también válidas; por ejemplo, la regla de la cadena

$$\frac{d\mathbf{r}[u(t), v(t)]}{dt} = \frac{\partial \mathbf{r}}{\partial u} \frac{du}{dt} + \frac{\partial \mathbf{r}}{\partial v} \frac{dv}{dt} = \mathbf{r}_u u' + \mathbf{r}_v v', \quad (7)$$

donde $\mathbf{r}_u$ y $\mathbf{r}_v$ son las derivadas parciales de la función vectorial $\mathbf{r}(u, v)$.

La longitud de una curva, como puede probarse, está expresada por la integral

$$s = \int \sqrt{x'^2(t) + y'^2(t) + z'^2(t)} dt.$$

De este modo, la diferencial de la longitud de una curva es igual a

$$ds = \sqrt{x'^2(t) + y'^2(t) + z'^2(t)} dt.$$

²⁷ Naturalmente, Gauss no utilizó notación vectorial, sino que definió las tres coordenadas x , y , z de las coordenadas de los puntos de la superficie separadamente como funciones de u y v . Los vectores, que se introdujeron a raíz de los trabajos de Hamilton y Grassmann, se utilizaron primero ampliamente en física, y sólo más tarde (en realidad, en el siglo xx) se convirtieron en el instrumento tradicional de la geometría analítica y diferencial.

Pero como $x'(t)$, $y'(t)$ y $z'(t)$ son las componentes del vector $dr/dt = \mathbf{r}'_t$, podemos escribir $ds = |\mathbf{r}'_t| dt$, donde $|\mathbf{r}'_t|$ denota la longitud del vector $\mathbf{r}'_t$. Para curvas sobre una superficie, obtenemos de (7)

$$ds = |\mathbf{r}_u u'_t + \mathbf{r}_v v'_t| dt.$$

Calculando el cuadrado de la longitud del vector del segundo miembro obtenemos, mediante las reglas del álgebra vectorial ²⁸,

$$ds^2 = [\mathbf{r}_u^2 u'^2 + 2\mathbf{r}_u \mathbf{r}_v u'_t v'_t + \mathbf{r}_v^2 v'^2] dt^2.$$

Pasando a diferenciales e introduciendo la notación

$$\mathbf{r}_u^2 = E(u, v), \quad \mathbf{r}_u \mathbf{r}_v = F(u, v), \quad \mathbf{r}_v^2 = G(u, v),$$

tenemos

$$ds^2 = E du^2 + 2F du dv + G dv^2.$$

Vemos que el cuadrado de la diferencial de la longitud de arco sobre una superficie es una forma cuadrática en las diferenciales du y dv con coeficientes que dependen del punto de la superficie. Esta forma se denomina la *primera forma cuadrática fundamental* de la superficie. Conocidos los coeficientes E , F y G de esta forma en cada punto de una superficie, podemos calcular la longitud de cualquier curva sobre ella mediante la fórmula

$$s = \int_{t_1}^{t_2} \sqrt{E u'^2 + 2F u'_t v'_t + G v'^2} dt,$$

de modo que su geometría intrínseca queda completamente determinada.

Veamos, a título de ejemplo, cómo se expresa el ángulo y el área mediante E , F y G . Sean dos curvas que parten de un punto dado, una de ellas dada por las ecuaciones $u = u_1(t)$, $v = v_1(t)$, y la otra, por $u = u_2(t)$, $v = v_2(t)$. Entonces las tangentes a estas curvas vienen dadas por los vectores

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{r}_u \frac{du_1}{dt} + \mathbf{r}_v \frac{dv_1}{dt}, \\ \mathbf{r}_2 &= \mathbf{r}_u \frac{du_2}{dt} + \mathbf{r}_v \frac{dv_2}{dt}. \end{aligned}$$

²⁸ El cuadrado de la longitud de un vector es el producto escalar del vector por sí mismo, y las reglas usuales para la multiplicación escalar (véase capítulo 3, §9) sirven para la supresión de corchetes.

El coseno del ángulo entre estos vectores es igual al producto escalar $\mathbf{r}_1 \mathbf{r}_2$ dividido por el producto de las longitudes $r_1 r_2$

$$\begin{aligned} \cos \alpha &= \frac{\mathbf{r}_1 \mathbf{r}_2}{r_1 r_2} = \\ &= \frac{r_u^2 \frac{du_1}{dt} \frac{du_2}{dt} + r_u r_v \left(\frac{du_1}{dt} \frac{dv_2}{dt} + \frac{du_2}{dt} \frac{dv_1}{dt} \right) + r_v^2 \frac{dv_1}{dt} \frac{dv_2}{dt}}{r_1 r_2}. \end{aligned}$$

Recordando que $r_u^2 = E$, $r_u r_v = F$, $r_v^2 = G$, obtenemos

$$\begin{aligned} \cos \alpha &= \\ &= \frac{E \frac{du_1}{dt} \frac{du_2}{dt} + F \left(\frac{du_1}{dt} \frac{dv_2}{dt} + \frac{du_2}{dt} \frac{dv_1}{dt} \right) + G \frac{dv_1}{dt} \frac{dv_2}{dt}}{\sqrt{E \left(\frac{du_1}{dt} \right)^2 + 2F \frac{du_1}{dt} \frac{dv_1}{dt} + G \left(\frac{dv_1}{dt} \right)^2} \sqrt{E \left(\frac{du_2}{dt} \right)^2 + 2F \frac{du_2}{dt} \frac{dv_2}{dt} + G \left(\frac{dv_2}{dt} \right)^2}}. \end{aligned}$$

Para obtener una fórmula del área, consideremos un rectángulo curvilíneo limitado por las curvas coordenadas $u = u_0$, $v = v_0$, $u = u_0 + \Delta u$, $v = v_0 + \Delta v$, y tomemos como una aproximación suya el paralelogramo situado en el plano tangente y limitado por los vectores $\mathbf{r}_u \Delta u$, $\mathbf{r}_v \Delta v$, tangentes a las curvas coordenadas (Fig. 46).

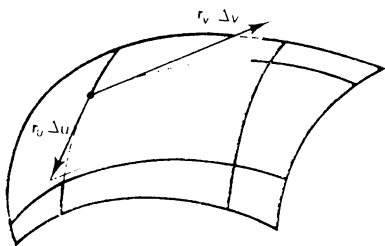


FIG. 46

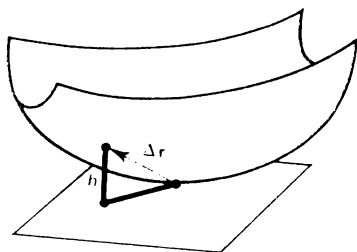


FIG. 47

El área de este paralelogramo es $\Delta s = |\mathbf{r}_u| |\mathbf{r}_v| \Delta u \Delta v \sin \phi$, donde ϕ es el ángulo entre $\mathbf{r}_u$ y $\mathbf{r}_v$. Como $\sin \phi = \sqrt{1 - \cos^2 \phi}$, se deduce que $\Delta s = |\mathbf{r}_u| |\mathbf{r}_v| \Delta u \Delta v \sqrt{1 - \cos^2 \phi} = \sqrt{r_u^2 r_v^2 - |\mathbf{r}_u|^2 |\mathbf{r}_v|^2 \cos^2 \phi} \Delta u \Delta v$.

Recordando que $r_u^2 = E$, $r_v^2 = G$, $|r_u| \cdot |r_v| \cos \phi = r_u r_v = F$, obtenemos $\Delta s = \sqrt{EG - F^2} \Delta u \Delta v$. Sumando las áreas de los paralelogramos y tomando el límite cuando $\Delta u, \Delta v \rightarrow 0$ obtenemos la fórmula para el área $S = \int \int_D \sqrt{EG - F^2} du dv$, donde la integración se toma sobre el dominio D de las variables u y v que describen el segmento dado de la superficie.

De esta manera, las coordenadas curvilíneas son muy convenientes para estudiar la geometría intrínseca de una superficie.

También se comprueba que la manera en que una superficie curvada está sumergida en el espacio circundante puede ser caracterizada mediante una cierta forma cuadrática en las diferenciales du, dv . Así, si $\mathbf{n}$ es un vector unitario normal a la superficie en el punto M , e $\Delta \mathbf{r}$ es el incremento del radio vector de la superficie cuando nos movemos desde este punto, entonces la desviación h de la superficie de su plano tangente (Fig. 47) es igual a $\mathbf{n} \Delta \mathbf{r}$. Desarrollando el incremento $\Delta \mathbf{r}$ por la fórmula de Taylor, obtenemos

$$h = \mathbf{n} d\mathbf{r} + \frac{1}{2} \mathbf{n} d^2\mathbf{r} + \epsilon(du^2 + dv^2),$$

donde $\epsilon \rightarrow 0$ cuando $\sqrt{du^2 + dv^2} \rightarrow 0$. Como el vector $d\mathbf{r}$ está en el plano tangente, tenemos $\mathbf{n} d\mathbf{r} = 0$. El último término, $\epsilon(du^2 + dv^2)$, es pequeño en comparación con los cuadrados de las diferenciales du y dv . Queda el término principal $\frac{1}{2} \mathbf{n} d^2\mathbf{r}$. Así, pues, dos veces la parte principal de h , es decir $\mathbf{n} d^2\mathbf{r}$, es una forma cuadrática con respecto a du y dv

$$\mathbf{n} d^2\mathbf{r} = \mathbf{n} r_{uu} du^2 + 2\mathbf{n} r_{uv} du dv + \mathbf{n} r_{vv} dv^2.$$

Esta forma describe el carácter de la desviación de la superficie a partir de su plano tangente. Se denomina la *segunda forma cuadrática fundamental* de la superficie. Sus coeficientes, que dependen de u y v , se suelen escribir así:

$$\mathbf{n} r_{uu} = L, \quad \mathbf{n} r_{uv} = M, \quad \mathbf{n} r_{vv} = N.$$

Conociendo la segunda forma cuadrática fundamental, podemos calcular la curvatura de cualquier curva sobre una superficie. Así, aplicando la fórmula $k = \lim_{l \rightarrow 0} 2h/l^2$, obtenemos el resultado de que la curvatura de la sección normal en la dirección correspondiente a la razón du/dv es igual a

$$k_n = \frac{\mathbf{n} d^2\mathbf{r}}{ds^2} = \frac{L du^2 + 2M du dv + N dv^2}{E du^2 + 2F du dv + G dv^2}.$$

Si la curva no es una sección normal, entonces, por el teorema de Meusnier, basta dividir la curvatura de la sección normal en la misma dirección por el coseno del ángulo entre la normal principal a la curva y la normal a la superficie.

La introducción de la segunda forma cuadrática fundamental proporciona un enfoque analítico para el estudio de cómo la superficie se curva en el espacio. En particular, se pueden deducir los teoremas de Euler y Meusnier, las expresiones para las curvaturas media y de Gauss, etc., de una forma puramente analítica.

El teorema de Peterson, mencionado anteriormente, prueba que las dos formas cuadráticas, tomadas conjuntamente, definen una superficie con exclusión de su posición en el espacio, de modo que el estudio de las propiedades de una superficie se reduce al de estas formas. En conclusión, observamos que los coeficientes de las dos formas cuadráticas no son independientes; la conexión mencionada anteriormente entre la geometría intrínseca de una superficie curvada y la forma en que está sumergida en el espacio viene expresada analíticamente por tres relaciones (las ecuaciones de Gauss-Codazzi) entre los coeficientes de la primera y la segunda formas cuadráticas fundamentales.

§5. Nuevos desarrollos en la teoría de curvas y superficies

Familias de curvas y superficies

Aun cuando la teoría básica de curvas y superficies estaba ya completada en gran medida hacia mediados del siglo pasado, ha seguido desarrollándose en nuevas direcciones que amplían enormemente la clase de figuras y propiedades investigadas en la geometría diferencial actual. Hay uno de estos desarrollos cuyos orígenes se remontan a los comienzos de la geometría diferencial, la teoría de «familias» o de colecciones continuas de curvas y superficies, aunque dicha teoría puede considerarse nueva en el sentido de que sus aspectos más profundos no fueron investigados hasta después de que la teoría básica de curvas y superficies estuvo totalmente desarrollada.

En general, una colección continua de figuras se denomina una *familia n -paramétrica* si cada figura de la colección está determinada por los valores de n parámetros y todas las cantidades que caracterizan la figura (respecto a su posición, forma, etc.) dependen de estos parámetros de una forma que es al menos continua. Desde el punto de vista de esta definición general, una curva puede ser consi-

derada como una familia uniparamétrica de puntos y una superficie como una familia biparamétrica de puntos. La colección de todos los círculos del plano es un ejemplo de una familia triparamétrica de curvas, ya que un círculo en el plano está determinado por tres parámetros: las dos coordenadas de su centro y su radio.

La cuestión más sencilla en la teoría de familias de curvas o superficies consiste en hallar la llamada *envolvente* de la familia. Una superficie se llama *envolvente* de una familia de superficies si en cada uno de sus puntos es tangente a una de las de la familia, y de esta manera tangente a cada una de ellas.

Por ejemplo, la envolvente de una familia de esferas de igual radio con centros sobre una recta dada será un cilindro (Fig. 48), y la envolvente de tales esferas con centros en todos los puntos de un plano dado

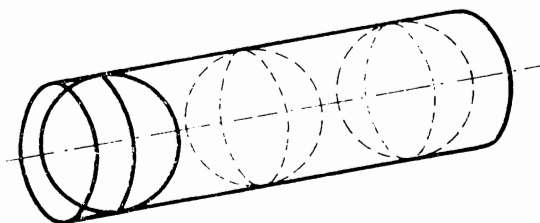


FIG. 48

consiste en dos planos paralelos. La envolvente de una familia de curvas se define análogamente. La figura 49 representa chorros de agua surgiendo de una fuente con distintos ángulos; en un plano cualquiera forman una familia de curvas que se pueden considerar aproximadamente como parábolas; su envolvente se destaca claramente como el contorno general de la cascada de agua. Naturalmente, no toda familia de curvas o superficies tiene una envolvente; por ejemplo una familia de rectas paralelas no posee ninguna. Hay un procedimiento general sencillo de calcular la envolvente de una familia; para una familia de curvas planas este método fue dado por Leibnitz.

Toda curva es evidentemente la envolvente de sus tangentes, del mismo modo que toda superficie es la envolvente de sus planos tangentes. Incidentalmente, este hecho aporta un nuevo método de definir una curva o una superficie dando la familia de sus rectas o planos tangentes. Para algunos problemas este método es el más conveniente.

Hablando en términos generales, los planos tangentes a una superficie son distintos en puntos diferentes, por lo que la familia de tangentes a la superficie es evidentemente biparamétrica. Sin embargo,

en ciertos casos —por ejemplo, en un cilindro— es uniparamétrica. Se puede probar que se verifica el siguiente teorema. Una familia uniparamétrica de planos tangentes se da solamente para aquellas superficies que son desarrollables en un plano, es decir, aquellas en las cuales un segmento suficientemente pequeño se puede deformar

en un segmento plano; estas son las superficies desarrollables indicadas en §4. Toda superficie analítica de este tipo está formada por segmentos de rectas y es, o cilíndrica (rectas paralelas), o cónica (rectas que pasan todas por un punto), o está formada por las tangentes a alguna curva alabeada.

La teoría de envolventes es particularmente útil en problemas de ingeniería, por ejemplo en la teoría de transmisiones. Consideremos dos engranajes *A* y *B*. Para estudiar el movimiento relativo entre ellos, podemos suponer que el en-

granaje *A* está fijo y el *B* se mueve en torno a él (Fig. 50). Entonces el perfil de un diente de *B*, al pasar por distintas posiciones, traza una familia de curvas en el plano de *A*, y el perfil de *A* debe ser en cada instante tangente a ellas, es decir, debe ser la envolvente de la familia. Naturalmente, esto no es un enunciado completo de la situación, ya que en una transmisión real este acoplamiento debe ser transferido de un par de dientes al siguiente; sin embargo, esta condición es la condición fundamental que debe satisfacer todo tipo de engranaje.

Como ya hemos dicho, la cuestión de las envolventes es uno de los problemas relativamente simples, resuelto hace mucho tiempo, de la teoría de familias de curvas y superficies. Esta teoría no es menos

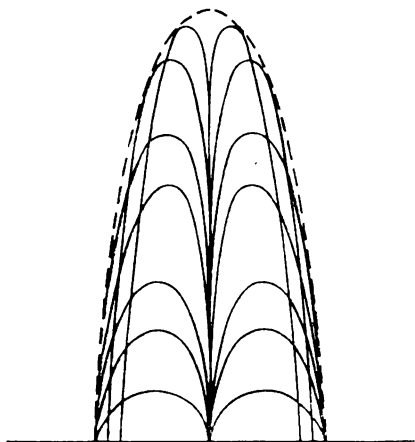


FIG. 49

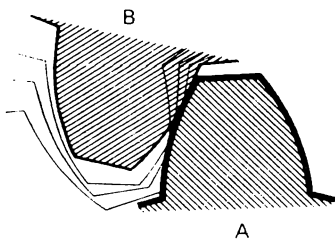


FIG. 50

rica en problemas interesantes que, digamos, la teoría de superficies misma. Especialmente bien desarrollada está la teoría de «congruencias» es decir, familias biparamétricas de distintas curvas (y en particular de rectas: las llamadas congruencias «de rectas»). En ella se utilizan esencialmente los mismos métodos que en la teoría de superficies.

La teoría de congruencias de rectas nació del trabajo de Monge, titulado «Sobre excavaciones y terraplenes», título que prueba que Monge emprendió la investigación con propósitos prácticos; la primera idea fue hallar la forma más conveniente de transportar tierra desde una excavación a un terraplén.

El desarrollo sistemático de la teoría de congruencias, que comenzó a mediados del siglo pasado, es debido en gran parte a su conexión con la óptica geométrica; el conjunto de rayos de luz en un medio homogéneo en cualquier instante constituye una congruencia de rectas.

Superficies no regulares y geometría «global»

La teoría de curvas y superficies (y familias de ellas), tal como había sido construida a finales del siglo pasado, se denomina usualmente geometría diferencial clásica y tiene los siguientes rasgos característicos.

En primer lugar, considera solamente curvas y superficies «suficientemente lisas» (es decir, regulares), o lo que es lo mismo, aquellas que están definidas por funciones que poseen un número suficiente de derivadas. Así, por ejemplo, superficies con cúspides o aristas, tales como superficies poliédricas o la superficie de un cono, o se excluyen de los argumentos o se consideran solamente sobre las partes en que son lisas.

En segundo lugar, la geometría diferencial clásica presta especial atención a las propiedades de segmentos suficientemente pequeños de curvas y superficies (geometría «local») y nunca considera propiedades de la superficie cerrada entera (geometría «global»).

Ejemplos típicos que ilustran la distinción entre geometría «local» y «global» nos los proporciona la deformación de superficies. Por ejemplo, ya en 1838 Minding probó que un segmento suficientemente pequeño de una esfera puede ser deformado; este es un teorema «local». Al mismo tiempo, expresó la conjetura de que la esfera entera no podía ser deformada. Este teorema fue probado por otros matemáticos en 1899. Incidentalmente, es fácil confirmar por la experiencia que una esfera de un material flexible aunque inextensible no se puede deformar. Por ejemplo, una pelota de ping-pong conserva su

forma perfectamente bien aunque el material de que esté hecha sea muy flexible. Otro ejemplo, mencionado en §4, es un cubo de hojalata; su rigidez global se debe a la presencia de un reborde curvado, pero cada una de sus partes, por separado, se pueden variar fácilmente de forma. Como vemos, hay una diferencia esencial entre propiedades «locales» y «globales».

Otros ejemplos característicos nos los proporciona la teoría de geodésicas, discutida en §4. Una geodésica «local», es decir, sobre un pequeño segmento de la superficie, es un camino mínimo, pero «globalmente» puede no ser así; por ejemplo, puede incluso ser una curva cerrada, como se apuntó anteriormente para los círculos máximos de una esfera.

El lector observará fácilmente que los teoremas sobre geodésicas formulados en §4 son básicamente «locales». Las cuestiones sobre el comportamiento de curvas geodésicas a lo largo de toda su longitud pertenecen a la geometría «global». Se sabe, por ejemplo, que, sobre una superficie regular, dos puntos suficientemente próximos se pueden unir por medio de una única geodésica que permanezca totalmente dentro de un cierto entorno pequeño de los dos puntos. Por el contrario, si consideramos geodésicas que durante su recorrido puedan desviarse tanto como queramos de ambos puntos, entonces, por un teorema debido a Morse, cualquier par de puntos sobre una superficie cerrada puede ser unido por una infinidad de geodésicas. De este modo, dos puntos A y B sobre la superficie lateral de un cilindro pueden ser unidos por geodésicas muy diferentes: basta considerar hélices que vayan de A a B , pero enrollándose alrededor del cilindro un número distinto de veces. El teorema de Poincaré sobre geodésicas cerradas, enunciado en la sección 5 del capítulo 17 y demostrado por Ljusternik y Šnirelman, también pertenece a la geometría «global».

Las demostraciones de estos teoremas, así como de otros muchos de geometría «global», resultaban inaccesibles con las herramientas usuales de la geometría diferencial clásica y requerían la invención de nuevos métodos.

Cuando estos problemas de geometría «global» atrajeron inevitablemente la atención de los matemáticos, la restricción a superficies regulares no pudo ser mantenida por más tiempo, aunque sólo fuese porque continuamente nos encontramos con superficies que no son regulares, sino que tienen curvatura discontinua; por ejemplo, lentes convexas con un borde afilado, etc. Además, hay muchas superficies analíticas que no se pueden extender de una manera natural sin que

adquieran «singularidades», en forma de aristas o cúspides, convirtiéndose por tanto en no regulares.

De este modo, un segmento de la superficie de un cono no puede ser extendido de manera natural sin conducirnos al vértice, una cúspide donde la regularidad de la superficie se destruye.

Este último resultado es solamente un caso particular del siguiente teorema: toda superficie desarrollable que no sea un cilindro conducirá, al extenderla de modo natural, a una arista (o una cúspide en el caso de un cono) más allá de la cual no puede ser prolongada sin que pierda su regularidad.

Así, pues, hay una profunda conexión entre el comportamiento global de una superficie y sus singularidades. Este es un motivo por el que la resolución de problemas «globales» y el estudio de superficies con «singularidades» deben realizarse juntos.

También el análisis tomó nuevos rumbos. Por ejemplo, la teoría cualitativa de ecuaciones diferenciales mencionada en la sección 7 del capítulo 5 estudia las propiedades de soluciones de una ecuación diferencial en su dominio total de definición, es decir, «globalmente», prestando particular atención a las «singularidades» (violaciones de la regularidad) y a los puntos singulares de la ecuación. Por otra parte, el análisis contemporáneo incluye el estudio de funciones no regulares, cosa que no se hace en el análisis clásico (capítulo 14), y de este modo proporciona a la geometría un nuevo medio de estudiar superficies más generales. Finalmente, en el cálculo de variaciones, que por lo general se utiliza para hallar curvas o superficies con alguna propiedad extrema, a veces sucede que la curva límite, para la cual se alcanza el extremo, no es regular. Para tales problemas es necesario que la clase de curvas o superficies en consideración sea cerrada (esto es, que incluya todas las curvas o superficies límites), hecho que necesariamente llevó al estudio de las curvas y superficies no regulares, o al menos al de las más sencillas. En una palabra, las nuevas direcciones tomadas por la geometría no se originaron aisladamente, sino en estrecha conexión con el desarrollo total de las matemáticas.

El interés por los problemas «globales» y por las superficies no regulares surgió hace unos cincuenta años y fue compartido por muchos matemáticos. El primer paso esencial lo dio Hermann Minkowski (1864-1909), quien preparó el fundamento de una extensa rama de la geometría, la teoría de cuerpos convexos. Una de las cuestiones que inició Minkowski en sus investigaciones fue el problema de retículos regulares, que está estrechamente conectado con la teoría de números y la cristalografía geométrica.

Un cuerpo se denomina *convexo* si por cada punto de su superficie puede hacerse pasar un plano que no corte al cuerpo, es decir, si en cualquier punto de su superficie el cuerpo puede quedar sobre un plano (Fig. 51). Un cuerpo convexo está determinado por su superficie exclusivamente, de modo que en general es indiferente hablar de cuerpos convexos o de superficies cerradas convexas. Los teoremas generales sobre cuerpos convexos se demuestran, por regla general, sin hacer hipótesis adicionales sobre la lisura o «regularidad» de sus

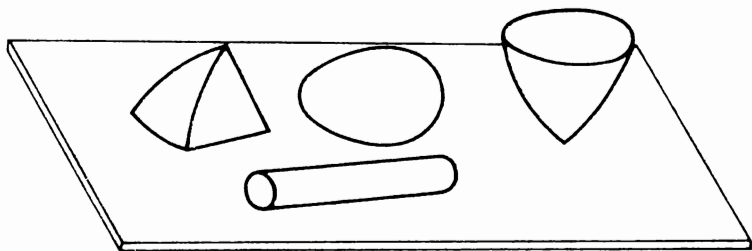


FIG. 51

superficies. Por tanto estos teoremas se refieren usualmente a la totalidad del cuerpo (o superficie) convexo, de modo que las restricciones de la geometría diferencial clásica se evitan automáticamente. Sin embargo, ambas teorías (la de cuerpos convexos y la de superficies no regulares) estuvieron al principio muy poco relacionadas entre sí; su combinación se produjo considerablemente más tarde.

A partir de 1940, A. D. Aleksandrov desarrolló la teoría de curvas y superficies generales, que incluye tanto las superficies regulares de la geometría diferencial clásica como superficies no lisas tales como poliedros, conjuntos convexos arbitrarios y otras. A pesar de la gran generalidad de esta teoría, está principalmente basada en conceptos y métodos geométricos intuitivos, aunque también hace uso esencial del análisis moderno. Uno de los métodos básicos de la teoría consiste en la aproximación de superficies generales mediante poliedros (superficies poliédricas). En su forma más simple, este artificio es conocido por cualquier escolar, aplicado, por ejemplo, al cálculo del área de la superficie lateral de un cilindro como límite de áreas de prismas. En muchos casos el método produce resultados fuertes que, o no pueden ser obtenidos por otro medio, o bien, si no pueden

ser probados por un procedimiento analítico, requieren la introducción de ideas complicadas. Su característica esencial consiste en el hecho de que el resultado se obtiene primeramente para poliedros, y después se extiende a superficies más generales por un proceso de paso al límite.

Uno de los comienzos de la teoría de superficies convexas generales fue el teorema sobre las condiciones bajo las cuales un desarrollo dado (Fig. 52) se puede pegar para formar un poliedro convexo. Este teorema, completamente elemental en su formulación, tiene una demostración no elemental y permite obtener corolarios para superficies convexas generales. El lector está, por supuesto, familiarizado con el ejercicio de ensamblar y pegar una superficie poliédrica: por ejemplo, la construcción de un cubo con el patrón en forma de cruz de la figura 52, o un cilindro a partir de un rectángulo y dos círculos.

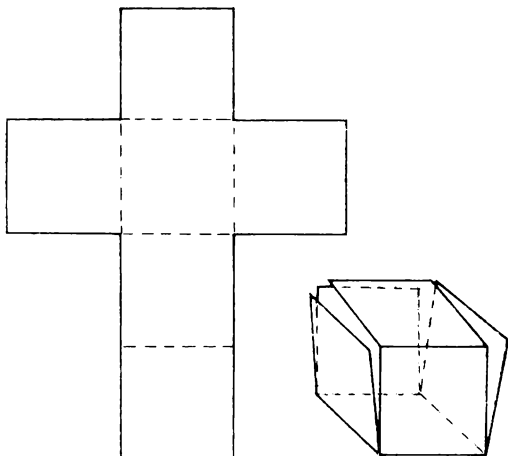


FIG. 52

El sencillo ejemplo de montar superficies a partir de sus trozos se convierte en un método general de «cortado y pegado», que ha producido resultados profundos en distintas cuestiones de la teoría de superficies y ha obtenido aplicaciones prácticas.

Profundos resultados en esta teoría fueron obtenidos por A. V. Pogorelov. En particular, probó que ninguna superficie convexa cerrada se puede deformar como un todo conservando su convexidad. Este resultado, conseguido en 1949, completa los esfuerzos de numerosos matemáticos que durante los cincuenta años precedentes habían tratado de demostrarlo, pero que solamente lo lograron a costa de introducir distintas hipótesis adicionales. Los resultados de Pogorelov, junto con el «método de pegar», no sólo proporcionó una solución completa del problema, sino que además aclaró del todo la cuestión

de la deformabilidad o no deformabilidad de superficies convexas cerradas o no cerradas. También establecieron una estrecha conexión entre la nueva teoría y la geometría diferencial «clásica».

De este modo se construyó una teoría de superficies que incluye la teoría clásica así como la teoría de poliedros, de superficies convexas arbitrarias y de superficies no convexas muy generales. La falta de espacio no nos permite discutir con detalle los resultados y problemas aún no resueltos de la teoría, aunque esto podría hacerse fácilmente, ya que son en su mayor parte de fácil visualización y, a pesar de la dificultad de una demostración exacta, no requieren conocimientos especiales.

En el §4, al hablar de la deformación de superficies, estábamos pensando en deformaciones de una superficie regular (curvada continuamente) que conservaba su regularidad. Pero en el teorema de Pogorelov, por el contrario, no se exige la regularidad ni de la superficie inicial ni de la deformada, aunque sí se impone a ambas el requisito de convexidad.

Es evidente que la deformación de una esfera, por ejemplo, sí es posible si permitimos roturas en la superficie y la violación de la convexidad. Basta quitar un segmento de la superficie y volverlo a colocar después de la deformación: esto es, por así decirlo, como empujar un segmento de la superficie al interior. Bastante más inesperado es el resultado obtenido recientemente por el matemático americano Nash y el holandés Kuiper. Estos demostraron que, si conservamos solamente la lisura de una superficie y permitimos la aparición de saltos bruscos en su curvatura (es decir, eliminamos algunas exigencias de continuidad, acotación o incluso la existencia de derivadas segundas de las funciones que definen la superficie), entonces es posible deformar la superficie como un todo con un alto grado de arbitrariedad. En particular, una esfera se puede deformar en una bola arbitrariamente pequeña formada por pliegues ondulados muy poco pronunciados. Una idea de una deformación de este tipo la proporciona la posibilidad, fácilmente imaginable, de arrugar casi de cualquier forma una funda esférica hecha de tela muy blanda. Por otra parte, una pequeña pelota de celuloide se comporta de una forma muy diferente. El material elástico de su superficie resiste no sólo a la dilatación, sino también al combamiento brusco, de manera que una pelota de este tipo es muy rígida.

Geometría diferencial de distintos grupos de transformaciones

Al principio de este siglo surgieron de la geometría diferencial clásica una serie de nuevos desarrollos basados en una idea general: el estudio de propiedades de curvas, superficies, y familias de unas y otras que permanecen invariantes bajo distintos tipos de transformaciones. La geometría diferencial clásica investigó propiedades invariantes bajo traslaciones; pero, por supuesto, no hay nada que impida considerar otras transformaciones geométricas. Por ejemplo, una *transformación proyectiva* es aquella bajo la cual las rectas siguen siendo rectas, y la geometría proyectiva, que existe desde hace mucho, estudia aquellas propiedades de las figuras que permanecen invariantes bajo transformaciones proyectivas. La geometría proyectiva ordinaria es similar, en los problemas que investiga, a la geometría elemental y analítica usual, mientras que la «geometría diferencial proyectiva» (la teoría de curvas, superficies y familias desarrollada al principio del presente siglo) es similar a la geometría diferencial clásica, excepto en que estudia propiedades que son invariantes bajo transformaciones proyectivas. Fundamentales en esta dirección fueron las contribuciones del americano Wilczynski, el italiano Fubini y el matemático checo Čech.

De la misma forma surgió la «geometría diferencial afín», que estudia las propiedades de curvas, superficies y familias invariantes bajo transformaciones afines, es decir, bajo transformaciones que no sólo conservan las rectas sino también el paralelismo. La labor del matemático alemán Blaschke y de sus discípulos convirtió esta rama de la geometría en una teoría general. Mencionemos también la «geometría conforme», en la que se estudian las propiedades de figuras invariantes bajo transformaciones que no alteran los ángulos entre curvas.

En general, las posibles «geometrías» son muy diversas en carácter, ya que, en esencia, cualquier grupo de transformaciones puede servir como base de una «geometría»; esta geometría estudiaría precisamente aquellas propiedades de las figuras que permanecen invariantes bajo las transformaciones del grupo. Este principio para la definición de geometrías será discutido más adelante en el capítulo 16.

Otras nuevas direcciones en geometría diferencial están siendo desarrolladas con éxito por los geómetras soviéticos S. P. Finikov, G. F. Laptev y otros. Pero en el presente libro no es posible mencionar todas y cada una de las distintas investigaciones que se llevan a cabo hoy día en las diferentes ramas de la geometría diferencial.

Capítulo 8

EL CALCULO DE VARIACIONES

§1. *Introducción*

Ejemplos de problemas variacionales

Podremos dar una descripción más clara del tipo general de problemas estudiados en el cálculo de variaciones ¹ si consideramos primero algunos problemas concretos.

1. *Curva de descenso más rápido*

El problema de la braquistócrona, o curva de descenso más rápido, fue históricamente el primer problema en el desarrollo del cálculo de variaciones.

Entre todas las curvas que unen los puntos M_1 y M_2 , se desea hallar aquélla a lo largo de la cual un punto matemático, moviéndose bajo la fuerza de la gravedad desde M_1 , sin velocidad inicial, llega al punto M_2 en el menor tiempo.

Para resolver este problema debemos considerar todas las posibles curvas que unen M_1 y M_2 . A una determinada curva l le corresponderá un valor determinado T del tiempo invertido para el descenso del punto material a lo largo de ella. El tiempo T dependerá de la elección de l , y de todas las curvas que unen M_1 con M_2 debemos hallar aquella a la que corresponda el menor valor de T .

¹ La procedencia del nombre «cálculo de variaciones» se explicará más tarde.

El problema de la braquistócrona puede ser expresado de la siguiente forma.

Tracemos un plano vertical que pase por los puntos M_1 y M_2 . La curva de más rápido descenso debe evidentemente estar en él, así que podemos restringirnos a curvas sobre dicho plano. Tomemos el punto M_1 como origen, el eje Ox horizontal y el eje Oy vertical y dirigido hacia abajo (Fig. 1). Las coordenadas del punto M_1 serán $(0, 0)$; las del punto M_2 las llamaremos (x_2, y_2) . Consideremos una curva arbitraria descrita por la ecuación

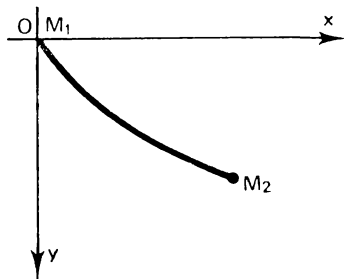


FIG. 1

$$y = f(x), \quad 0 \leq x \leq x_2, \quad (1)$$

donde f es una función continuamente diferenciable. Como la curva pasa por M_1 y M_2 , la función f en los extremos del segmento $[0, x_2]$ debe verificar la condición

$$f(0) = 0, \quad f(x_2) = y_2. \quad (2)$$

Si tomamos un punto arbitrario $M(x, y)$ sobre la curva, entonces la velocidad v de un punto material en este punto de la curva estará relacionada con la ordenada del punto por la ecuación física bien conocida

$$\frac{1}{2} v^2 = gy,$$

o bien:

$$v = \sqrt{2gy}.$$

El tiempo necesario para que un punto material recorra un elemento ds de arco de la curva tiene el valor

$$\frac{ds}{v} = \frac{\sqrt{1 + y'^2}}{\sqrt{2gy}} dx,$$

y por tanto el tiempo total de descenso del punto a lo largo de la curva desde M_1 a M_2 es igual a

$$T = \frac{1}{\sqrt{2g}} \int_0^{x_2} \frac{\sqrt{1 + y'^2}}{\sqrt{y}} dx. \quad (3)$$

Hallar la braquistócrona es equivalente a resolver el siguiente problema de mínimos: entre todas las posibles funciones (1) que verifiquen las condiciones (2), hallar la que corresponda al menor valor de la integral (3).

2. *La superficie de revolución de área mínima*

Entre las curvas que unen dos puntos de un plano, se desea hallar aquella cuyo arco, por rotación alrededor del eje Ox , engendra la superficie con menor área.

Denotemos los puntos dados por $M_1(x_1, y_1)$ y $M_2(x_2, y_2)$, y consideremos una curva arbitraria dada por la ecuación

$$y = f(x). \quad (4)$$

Si la curva pasa por M_1 y M_2 , la función f verificará la condición

$$f(x_1) = y_1, \quad f(x_2) = y_2. \quad (5)$$

Al girar alrededor del eje Ox esta curva describe una superficie con un área numéricamente igual al valor de la integral

$$S = 2\pi \int_{x_1}^{x_2} y \sqrt{1 + y'^2} dx. \quad (6)$$

Este valor depende de la elección de la curva, o lo que es equivalente, de la función $y = f(x)$. Entre todas las funciones (4) que verifiquen la condición (5) debemos hallar aquella que corresponde al menor valor de la integral (6).

3. *Deformación uniforme de una membrana*

Por membrana entendemos usualmente una superficie elástica que es plana en estado de reposo, se comba libremente, y sólo realiza un trabajo cuando se la somete a dilatación. Suponemos que la energía potencial de una membrana deformada es proporcional al incremento del área de su superficie.

Supongamos que en estado de reposo la membrana ocupa un dominio B del plano Oxy (Fig. 2). Deformamos el contorno de la membrana en una dirección perpendicular a Oxy y denotamos por $\phi(M)$ el desplazamiento del punto M del contorno. Entonces el interior de la membrana también se deforma, y deseamos hallar la posi-

ción de equilibrio de aquélla para una deformación conocida de su contorno.

Con un alto grado de exactitud podemos suponer que todos los puntos de la membrana se desplazan perpendicularmente al plano Oxy . Denotemos por $u(x, y)$ el desplazamiento del punto (x, y) . El área de la membrana en su posición desplazada será ²

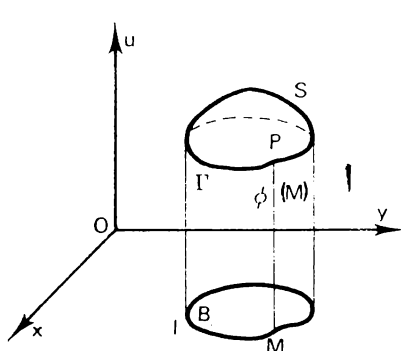


FIG. 2

$$\iint_B (1 + u_x^2 + u_y^2)^{1/2} dx dy.$$

Si las deformaciones de los elementos de la membrana son tan pequeñas que podemos ignorar potencias superiores de u_x y u_y , esta expresión para el área se puede reemplazar por una más simple:

$$\iint_B \left[1 + \frac{1}{2} (u_x^2 + u_y^2) \right] dx dy.$$

El incremento del área de la membrana es igual a

$$\frac{1}{2} \iint_B (u_x^2 + u_y^2) dx dy,$$

de modo que la energía potencial de la deformación tendrá el valor.

$$\frac{\mu}{2} \iint_B (u_x^2 + u_y^2) dx dy, \quad (7)$$

donde μ es una constante que depende de las propiedades elásticas de la membrana.

Como el desplazamiento de los puntos del borde de la membrana se supone conocido, la función $u(x, y)$ satisfará la condición

$$u|_l = \phi(M) \quad (8)$$

sobre el contorno del dominio B .

² Aquí, y a lo largo de todo el capítulo, usamos subíndices para denotar las variables respecto de las cuales se deriva parcialmente.

En la posición de equilibrio la energía potencial de la deformación debe tener el menor valor posible, con lo cual la función $u(x, y)$, que describe el desplazamiento de los puntos de la membrana, se halla resolviendo el siguiente problema matemático: entre todas las funciones $u(x, y)$ que son continuamente diferenciables en el dominio B y que verifican la condición (8) sobre el contorno, hallar aquella que da el menor valor de la integral (7).

Valores extremos de funcionales y el cálculo de variaciones

Estos ejemplos nos permiten formarnos una idea del tipo de problemas considerados; pero para determinar exactamente la situación del cálculo de variaciones en las matemáticas debemos familiarizarnos antes con algunos conceptos nuevos. Recordemos que uno de los conceptos básicos del análisis matemático es el de función. En el caso más simple el concepto de dependencia funcional se puede describir como sigue. Sea M un conjunto de números reales. Si cada número x del conjunto M se corresponde con un número y , decimos que sobre el conjunto M hay definida una función $y = f(x)$. El conjunto M se denomina con frecuencia el dominio de definición de la función.

El concepto de funcional es una generalización directa y natural del de función e incluye a ésta como un caso especial.

Sea M un conjunto de objetos de cualquier clase. La naturaleza de estos objetos no importa ahora. Pueden ser números, puntos de un espacio, curvas, funciones, superficies, estados o incluso movimientos de un sistema mecánico. Por brevedad los llamaremos elementos del conjunto M y los denotaremos por la letra x .

Si cada elemento x del conjunto M se corresponde con un número y , decimos que sobre el conjunto M hay definido un funcional $y = F(x)$.

Si M es un conjunto de números x , el funcional $y = F(x)$ será una función de un solo argumento. Cuando M es un conjunto de pares de números (x_1, x_2) o de puntos de un plano, el funcional será una función $y = F(x_1, x_2)$ de dos argumentos, etc.

Para el funcional $y = F(x)$ enunciemos el siguiente problema:

Entre todos los elementos x de M hallar aquel para el cual el funcional $y = F(x)$ toma el menor valor.

El problema de máximo de un funcional se formula de la misma manera.

Observemos que si cambiamos el signo en el funcional $F(x)$ y consideramos el $-F(x)$, el máximo (mínimo) de $F(x)$ se convierte en el mínimo (máximo) de $-F(x)$. Por tanto, no hay necesidad de

estudiar el máximo y el mínimo; en lo que sigue nos ocuparemos principalmente de mínimos de funcionales.

En el problema de la curva de descenso más rápido, el funcional cuyo mínimo buscamos será la integral (3), el tiempo de descenso de un punto material a lo largo de una curva. Este funcional estará definido sobre todas las posibles funciones (1) que verifiquen la condición (2).

En el problema de la posición de equilibrio de una membrana, el funcional es la energía potencial (7) de la membrana deformada, y debemos hallar su mínimo sobre el conjunto de funciones $u(x, y)$ que verifican la condición de contorno (8).

Todo funcional está determinado por dos factores: el conjunto M de elementos x sobre el cual está definido y la ley por la que a cada elemento x se le asocia un número, el valor del funcional. Los procedimientos de buscar el menor y mayor valores de un funcional dependerán indudablemente de las propiedades del conjunto M .

El cálculo de variaciones es un capítulo particular de la teoría de funcionales. En él consideramos funcionales definidos sobre un conjunto de funciones, y el problema consiste en la construcción de una teoría de valores extremos para tales funcionales.

Esta rama de las matemáticas llegó a ser particularmente importante después del descubrimiento de su conexión con muchas situaciones de la física y la mecánica. La razón de esta conexión se puede ver como sigue. Como ya aclararemos más adelante, para que una función proporcione un valor extremo de un funcional, es necesario que verifique una cierta ecuación diferencial. Por otra parte, como ya mencionamos en los capítulos dedicados a ecuaciones diferenciales, las cuantitativas de la mecánica y la física se escriben con frecuencia en forma de tales ecuaciones. Y resultó que muchas ecuaciones de este tipo también aparecían entre las ecuaciones diferenciales del cálculo de variaciones. De este modo fue posible considerar las ecuaciones de la mecánica y la física como condiciones extremas para funcionales convenientes y enunciar las leyes de la física en la forma de exigir un valor extremo, en particular un mínimo, para ciertas cantidades. Todo ello introdujo nuevos puntos de vista en la física y la mecánica, ya que ciertas leyes se podrían sustituir por enunciados equivalentes en términos de «principios de mínimo». Lo cual, a su vez, abrió la posibilidad de un nuevo método de resolución de problemas físicos, exacta o aproximadamente, buscando los mínimos de los correspondientes funcionales.

§2. Las ecuaciones diferenciales del cálculo de variaciones

La ecuación diferencial de Euler

El lector recordará que una condición necesaria para la existencia de un valor extremo de una función diferenciable f en el punto x es que la derivada f' sea igual a cero en este punto: $f'(x) = 0$; o lo que es lo mismo, que la diferencial de la función sea igual a cero: $df = f'(x) dx = 0$.

Nuestra meta inmediata será hallar una condición análoga en el cálculo de variaciones, es decir, obtener una condición necesaria que deba verificar una función para que proporcione un valor extremo del funcional.

Mostraremos que una función tal debè verificar una cierta ecuación diferencial. La forma de la ecuación dependerá del tipo de funcional que consideremos. Empecemos con la integral más sencilla del cálculo de variaciones, es decir, un funcional con la siguiente representación integral:

$$I(y) = \int_{x_1}^{x_2} F(x, y, y') dx. \quad (9)$$

La función F que aparece bajo el signo integral depende de tres argumentos (x, y, y') . Supondremos que está definida y que es dos veces derivable con continuidad respecto a y' para todos los valores de este argumento, y con respecto a x e y en algún dominio B del plano Oxy . En lo que sigue se supone que siempre permanecemos en el interior de este dominio.

Está claro que y es una función de x

$$y = y(x), \quad (10)$$

continuamente diferenciable sobre el segmento $x_1 \leq x \leq x_2$, y que y' es su derivada.

Geométricamente, la función $y(x)$ se puede representar en el plano Oxy por una curva l sobre el intervalo $[x_1, x_2]$ (Fig. 3).

La integral (9) es una generalización de las integrales (3) y (6) que ya encontramos en el problema de la curva de descenso más

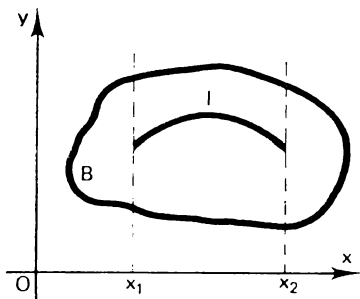


FIG. 3

rápido y en el de la superficie de revolución de menor área. Su valor depende de la elección de la función $y(x)$ o, en otras palabras, de la curva I , y el problema de su valor mínimo se interpreta como sigue:

Sea un conjunto M de funciones (10) (curvas I); entre éstas debemos hallar aquella función (curva I) para la cual la integral $I(y)$ tiene el menor valor.

Antes de nada es preciso definir el conjunto M de funciones para las que consideraremos el valor de la integral (9). En el cálculo de variaciones las funciones de este conjunto se suelen denominar «admisibles». Consideremos el problema con valores fijos en el contorno. El conjunto de funciones admisibles está determinado aquí por las dos condiciones siguientes:

1. $y(x)$ es continuamente diferenciable sobre el segmento $[x_1, x_2]$.
2. En los extremos del segmento $y(x)$ toma valores prefijados

$$y(x_1) = y_1, \quad y(x_2) = y_2. \quad (11)$$

Por lo demás, la función $y(x)$ puede ser completamente arbitraria. En lenguaje geométrico, consideramos todas las posibles curvas lisas que, situadas sobre el intervalo $[x_1, x_2]$, pasan por los dos puntos $A(x_1, y_1)$ y $B(x_2, y_2)$ y pueden representarse mediante la ecuación (10). Supondremos que la función que da el mínimo de la integral existe, y se la llamará $y(x)$.

Los siguientes argumentos, sencillos e ingeniosos a la par que aplicables con frecuencia en el cálculo de variaciones, conducen a una forma particularmente simple de la condición necesaria que debe verificar $y(x)$. En esencia, dichos argumentos nos permiten reducir el problema del mínimo de la integral (9) al problema del mínimo de una función.

Consideremos la familia de funciones dependiente de un parámetro numérico α ,

$$\tilde{y}(x) = y(x) + \alpha\eta(x). \quad (12)$$

Para que $\tilde{y}(x)$ sea una función admisible para un α arbitrario, debemos suponer que $\eta(x)$ es continuamente diferenciable y se anula en los extremos del intervalo $[x_1, x_2]$.

$$\eta(x_1) = \eta(x_2) = 0. \quad (13)$$

La integral (9) calculada para $\bar{y}$ será una función del parámetro α

$$I(\bar{y}) = \int_{x_1}^{x_2} F(x, y + \alpha\eta, y' + \alpha\eta') dx = \Phi(\alpha).^3$$

Como $y(x)$ da un mínimo para el valor de la integral, la función $\Phi(\alpha)$ debe tener un mínimo para $\alpha = 0$, de modo que su derivada en este punto debe anularse

$$\Phi'(0) = \int_{x_1}^{x_2} [F_y(x, y, y') \eta + F_{y'}(x, y, y') \eta'] dx = 0. \quad (14)$$

Esta última ecuación se debe verificar para toda función continuamente diferenciable $\eta(x)$ que se anule en los extremos del segmento $[x_1, x_2]$. Para obtener el resultado que se deduce de esto, es conveniente transformar el segundo miembro de la condición (14) mediante integración por partes

$$\int_{x_1}^{x_2} F_{y'} \eta' dx = - \int_{x_1}^{x_2} \eta \frac{d}{dx} F_{y'} dx,$$

de donde la condición (14) toma la nueva forma

$$\Phi'(0) = \int_{x_1}^{x_2} \left(F_y - \frac{d}{dx} F_{y'} \right) \eta dx = 0. \quad (15)$$

Se puede probar que el siguiente lema es válido.

Supongamos que se verifican las dos siguientes condiciones:

1. La función $f(x)$ es continua en el intervalo $[a, b]$.
2. La función $\eta(x)$ es continuamente diferenciable en el intervalo $[a, b]$ y se anula en los extremos del mismo.

Entonces, si para una función arbitraria $\eta(x)$ la integral $\int_a^b f(x) \eta(x) dx$ es igual a cero, se deduce que $f(x) \equiv 0$.

Para ello supongamos que en algún punto c la función f es distinta

³ La diferencia $\bar{y} - y = \alpha\eta$ se denomina *variación* (cambio) de la función y y se denota por δy , y la diferencia $I(\bar{y}) - I(y)$ se denomina *variación total* de la integral (9). De aquí el nombre de cálculo de variaciones.

de cero y probemos que entonces existe necesariamente una función $\eta(x)$ para la cual $\int_a^b f(x) \eta(x) dx \neq 0$, en contradicción con la condición del lema.

Como $f(c) \neq 0$ y f es continua, debe existir un entorno $[\alpha, \beta]$ de c en el cual f sea distinta de cero y, por tanto, tenga signo constante en todo él.

Siempre es posible construir una función $\eta(x)$ que sea continuamente diferenciable sobre $[a, b]$, positiva en $[\alpha, \beta]$, e igual a cero en el exterior de $[\alpha, \beta]$ (Fig. 4).

Una función $\eta(x)$ tal se puede definir, por ejemplo, por las ecuaciones

$$\eta(x) = \begin{cases} 0 & \text{sobre } [a, \alpha], \\ (x - \alpha)^2(\beta - x)^2 & \text{sobre } [\alpha, \beta], \\ 0 & \text{sobre } [\beta, b]. \end{cases}$$

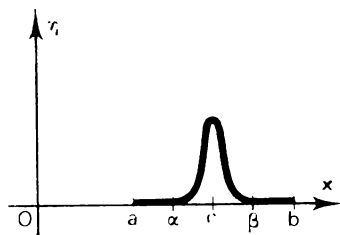


FIG. 4

Pero para tal función $\eta(x)$ tenemos:

$$\int_a^b f \eta dx = \int_a^b f \eta dx.$$

La segunda de estas integrales no puede ser igual a cero, ya que, en el interior del intervalo de integración, el producto $f\eta$ es distinto de cero y nunca cambia de signo.

Como la ecuación (15) se debe verificar para toda $\eta(x)$ que sea continuamente diferenciable y se anule en los extremos del segmento $[x_1, x_2]$, podemos asegurar, apoyándonos en el lema, que esto sólo puede ocurrir en el caso en que

$$F_y - \frac{d}{dx} F_{y'} = 0, \quad (16)$$

o, calculando la derivada con respecto a x ,

$$F_y(x, y, y') - F_{xy'}(x, y, y') - F_{yy'}(x, y, y')y' - F_{y'y'}(x, y, y')y'' = 0. \quad (17)$$

Esta es una ecuación diferencial de segundo orden con respecto a la función y . Se denomina *ecuación de Euler*.

Podemos enunciar la siguiente conclusión.

Si una función $y(x)$ minimiza la integral $I(y)$, entonces debe verificar la ecuación diferencial de Euler (17). En el cálculo de variaciones, este último enunciado tiene un sentido completamente análogo

a la condición necesaria $df = 0$ en la teoría de valores extremos de funciones. Ello nos conduce inmediatamente a excluir todas las funciones admisibles que no verifiquen esta condición, ya que para ellas la integral no puede tener un mínimo; de este modo, el conjunto de funciones admisibles que necesitamos estudiar queda reducido drásticamente.

Las soluciones de la ecuación (17) tienen la propiedad de que para ellas la derivada $[(d/da)I(y + a\eta)]_{a=0}$ se anula para $\eta(x)$ arbitraria, por lo que tienen un significado análogo a los puntos estacionarios de una función. Así, a menudo se dice que para las soluciones de (17) la integral $I(y)$ tiene un valor estacionario.

En nuestro problema con valores de contorno fijos, no necesitamos hallar todas las soluciones de la ecuación de Euler, sino sólo aquellas que toman los valores y_1, y_2 en los puntos x_1, x_2 .

Volvamos nuestra atención al hecho de que la ecuación de Euler es de segundo orden. Su solución general contendrá dos constantes arbitrarias:

$$y = \phi(x, C_1, C_2).$$

Las dos constantes deben determinarse de modo que la curva integral pase por los puntos A y B , de forma que tenemos dos ecuaciones para hallar las constantes C_1 y C_2 :

$$\phi(x, C_1, C_2) = y_1, \quad \phi(x_2, C_1, C_2) = y_2.$$

En muchos casos este sistema tiene solamente una solución, y entonces existirá sólo una curva integral que pase por A y B .

La búsqueda de soluciones que den un mínimo de esta integral se reduce, por tanto, a la resolución del siguiente problema de contorno de ecuaciones diferenciales: en el intervalo $[x_1, x_2]$ hallar aquellas soluciones de la ecuación (17) que tomen los valores y_1, y_2 en los extremos del mismo.

Con frecuencia, este último problema se puede resolver utilizando métodos conocidos de la teoría de ecuaciones diferenciales.

Volvemos a hacer hincapié en que la solución de tal problema de contorno puede proporcionar sólo un *posible* mínimo, y que es necesario comprobar si da o no un mínimo para la integral. Pero en casos particulares, especialmente en aquellos que surgen en las aplicaciones, la ecuación de Euler resuelve completamente el problema de hallar el mínimo de la integral. Supongamos que sabemos inicialmente que existe una función que da un mínimo para la integral, y supongamos,

por otra parte, que la ecuación de Euler (17) tiene una sola solución que verifique las condiciones de contorno (11). Entonces sólo una de las curvas admisibles puede ser un posible mínimo, y podemos estar seguros, en estas circunstancias, de que la solución hallada de la ecuación (17) da realmente un mínimo para la integral.

Ejemplo

Quedó ya establecido que el problema de la curva de descenso más rápido se puede reducir a hallar el mínimo de la integral

$$I(y) = \int_0^{x_2} \frac{\sqrt{1 + y'^2}}{\sqrt{y}} dx$$

entre el conjunto de funciones que verifiquen las condiciones de contorno

$$y(0) = 0, \quad y(x_2) = y_2.$$

En este problema

$$F = \frac{\sqrt{1 + y'^2}}{\sqrt{y}}.$$

La ecuación de Euler tiene la forma

$$-\frac{1}{2} y^{-3/2} \sqrt{1 + y'^2} - \frac{d}{dx} \left[y^{-1/2} \cdot \frac{y'}{\sqrt{1 + y'^2}} \right] = 0.$$

Después de operar, toma la forma

$$\frac{2y''}{1 + y'^2} = -\frac{1}{y}.$$

Multiplicando ambos miembros de la ecuación por y' e integrando, obtenemos

$$\ln(1 + y'^2) = -\ln y + \ln k,$$

o bien

$$y'^2 = \frac{k}{y} - 1,$$

$$\int \frac{y}{k - y} dy = \pm dx.$$

Haciendo el cambio

$$y = \frac{k}{2}(1 - \cos u), \quad dy = \frac{k}{2} \sin u \, du,$$

hallamos, después de sustituir y simplificar,

$$\frac{k}{2}(1 - \cos u) \, du = \pm dx,$$

de donde, integrando, obtenemos: $x = \pm \frac{k}{2}(u - \sin u) + C$. Como la curva debe pasar por el origen, se deduce que debemos poner $C = 0$.

De esta manera vemos que la braquistócrona es la cicloide

$$x = \frac{k}{2}(u - \sin u), \quad y = \frac{k}{2}(1 - \cos u).$$

La constante k se halla haciendo uso de la condición de que esta curva pase por el punto $M_2(x_2, y_2)$.

Funcionales dependientes de varias funciones

El funcional más sencillo del cálculo de variaciones (17) dependía sólo de una función. En las aplicaciones, tales funcionales se darán en aquellos casos en que los objetos (o su comportamiento) estén determinados por una sola dependencia funcional. Por ejemplo, una curva plana está definida por la dependencia de la ordenada de un punto con respecto a su abscisa; el movimiento de un punto material a lo largo de un eje está determinado por la dependencia de sus coordenadas con respecto al tiempo, etc.

Pero a menudo hay que tratar con objetos que no pueden determinarse de una manera tan simple. Para definir una curva alabeada debemos conocer la dependencia funcional de dos de sus coordenadas con respecto a la tercera. El movimiento de un punto en el espacio está determinado por la dependencia de sus tres coordenadas con respecto al tiempo, etc. El estudio de estos objetos más complicados nos lleva a problemas variacionales con diversas funciones que varían.

Nos restringiremos a los casos en que el funcional dependa de dos funciones $y(x)$ y $z(x)$, ya que el caso de un mayor número de funciones no se diferencia en principio de éste.

Consideremos el siguiente problema. Los pares admisibles de funciones están definidos por las siguientes condiciones:

1. Las funciones

$$y = y(x), \quad z = z(x) \quad (18)$$

son continuamente diferenciables en el segmento $[x_1, x_2]$.

2. En los extremos del segmento estas funciones toman valores prefijados

$$\begin{aligned} y(x_1) &= y_1, & y(x_2) &= y_2 \\ z(x_1) &= z_1, & z(x_2) &= z_2. \end{aligned} \quad (19)$$

Entre todos los posibles pares de funciones $y(x)$ y $z(x)$, debemos hallar aquel que da el menor valor de la integral

$$I(y, z) = \int_{x_1}^{x_2} F(x, y, z, y', z') dx. \quad (20)$$

En el espacio tridimensional x, y, z , cada par de funciones admisibles corresponderá a una curva l , definida por las ecuaciones (18) y que pasa por los puntos

$$M_1(x_1, y_1, z_1), \quad M_2(x_2, y_2, z_2).$$

Debemos hallar el mínimo de la integral (20) sobre el conjunto de tales curvas.

Supongamos que el par de funciones que da el mínimo de la integral (20) existe, y llamemos a estas funciones $y(x)$ y $z(x)$. Junto con ellas consideremos un segundo par de funciones

$$\bar{y} = y + \alpha \eta(x), \quad \bar{z} = z + \alpha \zeta(x),$$

donde $\eta(x)$ y $\zeta(x)$ son funciones continuamente diferenciables que se anulan en los extremos del segmento $[x_1, x_2]$; $\bar{y}$, $\bar{z}$ serán también admisibles, y para $\alpha = 0$ coincidirán con las funciones y, z . Sustituyéndolas en (20), tenemos:

$$I(\bar{y}, \bar{z}) = \int_{x_1}^{x_2} F(x, y + \alpha \eta, z + \alpha \zeta, y' + \alpha \eta', z' + \alpha \zeta') dx = \Phi(\alpha).$$

La integral así deducida será una función de α . Como $\bar{y}$ y $\bar{z}$ coinciden con y y z cuando $\alpha = 0$, la función $\Phi(\alpha)$ debe tener un mínimo para $\alpha = 0$. Pero en un punto mínimo la derivada de Φ debe anularse

$$\Phi'(0) = 0.$$

Calculando la derivada da

$$\int_{x_1}^{x_2} (F_y \cdot \eta + F_z \cdot \zeta + F_{y'} \cdot \eta' + F_{z'} \cdot \zeta') dx = 0,$$

o, si los términos en η' y ζ' se integran por partes,

$$\int_{x_1}^{x_2} \left[\left(F_y - \frac{d}{dx} F_{y'} \right) \eta(x) + \left(F_z - \frac{d}{dx} F_{z'} \right) \zeta(x) \right] dx = 0.$$

Esta última ecuación debe verificarse para dos funciones continuamente diferenciables cualesquiera $\eta(x)$ y $\zeta(x)$ que se anulen en los extremos del intervalo. Por consiguiente, en virtud del lema básico probado anteriormente, se deben satisfacer las dos siguientes condiciones:

$$\begin{aligned} F_y - \frac{d}{dx} F_{y'} &= 0 \\ F_z - \frac{d}{dx} F_{z'} &= 0 \end{aligned} \tag{21}$$

Por tanto, si las funciones y , z dan un mínimo para la integral (20), deben verificar el sistema de ecuaciones diferenciales de Euler (21).

Este resultado nos permite una vez más reemplazar un problema variacional para el mínimo de la integral (20) por un problema de contorno en la teoría de ecuaciones diferenciales: en el intervalo $[x_1, x_2]$ debemos hallar aquellas soluciones y , z del sistema de ecuaciones diferenciales (21) que verifiquen las condiciones de contorno (19).

Como en el caso precedente, esto nos abre un posible camino para la resolución del problema minimal.

Como ejemplo de una aplicación del sistema de Euler (21), consideremos el principio variacional de Ostrogradski-Hamilton en la

mecánica newtoniana. Nos restringiremos a la forma más simple de este principio.

Consideremos un cuerpo material de masa m y supongamos que podemos ignorar las dimensiones y la forma del mismo, de modo que podemos imaginarlo como un punto material.

Supongamos que el punto se mueve desde la posición $M_1(x_1, y_1, z_1)$ en el instante t_1 hasta la posición $M_2(x_2, y_2, z_2)$ en el instante t_2 . Supongamos también que el movimiento tiene lugar bajo las leyes de la mecánica newtoniana y está causado por la aplicación de una fuerza $F(x, y, z, t)$ que depende de la posición del punto y del tiempo t y que posee una función potencial $U(x, y, z, t)$. La última condición significa lo siguiente: las componentes F_x, F_y, F_z de la fuerza F según los ejes de coordenadas, son las derivadas parciales de una función U con respecto a las correspondientes coordenadas

$$F_x = \frac{\partial U}{\partial x}, \quad F_y = \frac{\partial U}{\partial y}, \quad F_z = \frac{\partial U}{\partial z}.$$

Supongamos que el movimiento es libre, es decir, no sujeto a ningún tipo de ligaduras ⁴.

Las ecuaciones del movimiento de Newton son

$$m \frac{d^2 x}{dt^2} = \frac{\partial U}{\partial x}, \quad m \frac{d^2 y}{dt^2} = \frac{\partial U}{\partial y}, \quad m \frac{d^2 z}{dt^2} = \frac{\partial U}{\partial z}.$$

Si el punto obedece las leyes de la mecánica newtoniana, se mueve de una manera completamente determinada. Pero, junto con estos «movimientos newtonianos» del punto, consideremos otros (no newtonianos), que por brevedad llamaremos «admisibles» y que estarán determinados por dos condiciones solamente: que en el instante t_1 el punto esté en la posición M_1 , y en el instante t_2 esté en la posición M_2 .

¿Cómo podemos distinguir el «movimiento newtoniano» del punto de los otros movimientos «admisibles»? Tal posibilidad viene dada por el principio de Ostrogradski-Hamilton.

Introduzcamos la energía cinética del punto

$$T = \frac{1}{2} m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)$$

⁴ Esto no es esencial para el principio de Ostrogradski-Hamilton: podemos imponer cualquier ligadura que deseemos al sistema mecánico, incluso ligaduras no estacionarias, con la única condición de que sean holónomas, es decir, que se puedan escribir mediante ecuaciones en las que no figuren derivadas de las coordenadas con respecto al tiempo.

y formemos la llamada integral de acción

$$I = \int_{t_1}^{t_2} (T + U) dt.$$

El principio dice: el «movimiento newtoniano» del punto se distingue de todos sus movimientos «admisibles» por el hecho de que da, para la integral de acción, un valor estacionario.

La integral de acción I depende de tres funciones: $x(t)$, $y(t)$, $z(t)$.

Como para todos los movimientos las posiciones inicial y final del punto son idénticas, los valores de contorno de estas funciones son fijos. Se trata de un problema variacional para tres funciones que varían con valores fijos en los extremos del intervalo $[t_1, t_2]$.

Previamente hemos convenido en decir que la integral (17) tiene un valor estacionario para toda curva que sea una curva integral de la ecuación de Euler. En nuestro problema integramos una función

$$F = T + U = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + U(x, y, z, t),$$

que depende de tres funciones, de modo que para un valor estacionario de la integral se debe verificar el sistema de tres ecuaciones diferenciales

$$F_x - \frac{d}{dt} F_{\dot{x}} = 0,$$

$$F_y - \frac{d}{dt} F_{\dot{y}} = 0,$$

$$F_z - \frac{d}{dt} F_{\dot{z}} = 0.$$

Como $F_x = \partial U / \partial x$, $F_{\dot{x}} = m\dot{x}$, ..., el sistema de ecuaciones de Euler es idéntico a las ecuaciones del movimiento en mecánica newtoniana, lo que proporciona una comprobación del principio de Ostrogradski-Hamilton.

El problema del mínimo para una integral múltiple

El último problema del cálculo de variaciones sobre el que queremos llamar la atención del lector es el de minimizar una integral múltiple. Como los hechos relacionados con la resolución de tales

problemas son análogos para integrales de cualquier multiplicidad, nos limitaremos al caso más sencillo, el de integrales dobles.

Sea B un dominio en el plano Oxy , limitado por el contorno l . El conjunto de funciones admisibles está determinado por las condiciones:

1. $u(x, y)$ es continuamente diferenciable sobre el dominio B .
2. Sobre l , la función u toma los valores prefijados

$$u|_l = f(M). \quad (22)$$

Entre todas las funciones debemos hallar aquella que dé un valor mínimo para la integral

$$I(u) = \iint_B F(x, y, u, u_x, u_y) dx dy. \quad (23)$$

Los valores de contorno dados (22) para la función u en el espacio (x, y, u) determinan una curva alabeada conocida Γ , situada por encima de l (véase figura 2, capítulo 7).

Consideremos todas las posibles superficies S que contengan a Γ y que estén situadas por encima de B . Entre éstas queremos hallar aquella para la que la integral (23) es mínima.

Como antes, suponemos la existencia de la función minimizante y la denotamos por u . Al mismo tiempo consideramos otra función

$$\bar{u} = u + \alpha\eta(x, y),$$

donde $\eta(x, y)$ es una función continuamente diferenciable que se anula sobre l . Entonces la función

$$I(\bar{u}) = \iint_B F(x, y, u + \alpha\eta, u_x + \alpha\eta_x, u_y + \alpha\eta_y) dx dy = \Phi(\alpha)$$

debe tener un mínimo para $\alpha = 0$. En este caso su derivada primera debe ser igual a cero para $\alpha = 0$

$$\Phi'(0) = 0,$$

o bien

$$\iint_B (F_u\eta + F_{u_x}\eta_x + F_{u_y}\eta_y) dx dy = 0. \quad (24)$$

Transformamos los dos últimos términos por la fórmula de Ostrogradski

$$\begin{aligned} \iint_B (F_{u_x} \eta_x + F_{u_y} \eta_y) dx dy = \\ \iint_B \left(\frac{\partial}{\partial x} (F_{u_x} \eta) + \frac{\partial}{\partial y} (F_{u_y} \eta) \right) dx dy - \iint_B \left(\frac{\partial}{\partial x} F_{u_x} + \frac{\partial}{\partial y} F_{u_y} \right) \eta dx dy = \\ = \int_l [F_{u_x} \cos(n, x) + F_{u_y} \cos(n, y)] \eta ds - \\ - \iint_B \left(\frac{\partial}{\partial x} F_{u_x} + \frac{\partial}{\partial y} F_{u_y} \right) \eta dx dy. \end{aligned}$$

La integral de contorno a lo largo de l debe anularse, ya que sobre el contorno l la función η es igual a cero, por lo que la condición (24) puede ponerse en la forma

$$\iint_B \left(F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} \right) \eta dx dy = 0.$$

Esta ecuación se debe verificar para toda función η que sea continuamente diferenciable y se anule sobre el contorno l .

Podemos concluir, como antes, que en todos los puntos del dominio B la ecuación

$$F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} = 0 \quad (25)$$

debe ser satisfecha.

Así, si la función u da un mínimo para la integral (23), debe verificar la ecuación en derivadas parciales (25).

Como en todos los problemas precedentes, hemos establecido aquí una conexión entre un problema variacional de minimizar una integral y un problema de contorno para una ecuación diferencial (en este caso en derivadas parciales).

Ejemplo

El desplazamiento $u(x, y)$ de los puntos de una membrana con un contorno deformado se halla a partir de la condición de mínimo de la energía potencial

$$\frac{\mu}{2} \iint_B (u_x^2 + u_y^2) dx dy$$

para los valores de contorno dados $u|_l = \phi$.

Omitiendo, por simplicidad, el factor constante μ , podemos poner

$$F = \frac{1}{2} (u_x^2 + u_y^2),$$

de modo que la ecuación (25) toma la forma

$$-\frac{\partial}{\partial x} u_x - \frac{\partial}{\partial y} u_y = 0,$$

o

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

De este modo el problema de determinar el desplazamiento de los puntos de una membrana ha sido reducido al de hallar una función armónica u con valores conocidos sobre el contorno del dominio (véase capítulo 6, §3).

§3. *Métodos de resolución aproximada de problemas en el cálculo de variaciones*

Concluimos el presente capítulo con una indicación acerca de las ideas implicadas en alguno de los métodos de aproximación del cálculo de variaciones.

Para concretar, discutiremos el funcional más sencillo:

$$I(y) = \int_{x_1}^{x_2} F(x, y, y') dx$$

para valores de contorno fijos de las funciones admisibles.

Sea $y(x)$ una solución exacta del problema de minimizar I , y $m = I(y)$ el correspondiente valor mínimo de la integral. Parece

natural que, si determinamos una función admisible $\bar{y}$ para la cual el valor de la integral $I(\bar{y})$ sea muy próximo a m , podamos suponer que $\bar{y}$ se diferenciará también muy poco de la solución exacta y . Además, si podemos construir una sucesión de funciones admisibles $\bar{y}_1, \bar{y}_2, \dots$ para la cual $I(\bar{y}_n) \rightarrow m$, podemos esperar que converja en un sentido u otro a la solución y , de modo que el cálculo de $\bar{y}_n$ con un subíndice suficientemente grande nos permitirá hallar la solución con el grado deseado de exactitud.

Según hagamos la elección de la «sucesión minimizante» $\bar{y}_n (n = 1, 2, \dots)$, tendremos uno u otro de los distintos métodos de aproximación del cálculo de variaciones.

Históricamente, el primero de ellos fue el método de las poligonales, o método de Euler. Descomponemos el intervalo $[x_1, x_2]$ en varios segmentos. Por ejemplo, si los elegimos de igual longitud, los puntos de división serán

$$x_1, x_1 + h, x_1 + 2h, \dots, x_1 + nh = x_2, h = \frac{x_2 - x_1}{n}$$

Construimos ahora la poligonal p_{n-1} , con vértices situados sobre los puntos de división. Las ordenadas de los vértices las denotamos por

$$b_0, b_1, b_2, \dots, b_{n-1}, b_n,$$

y exigimos que esta poligonal empiece y termine en los mismos puntos que las curvas admisibles, de modo que $b_0 = y_1$ y $b_n = y_2$. La poligonal estará determinada entonces por las ordenadas

$$b_1, b_2, \dots, b_{n-1}.$$

La cuestión ahora es cómo elegir la poligonal p_{n-1} (es decir, las ordenadas b_i de sus vértices) para que sea lo más próxima posible a la solución exacta del problema.

Para lograr este propósito es natural proceder como sigue. Calculamos la integral I para la poligonal. Su valor dependerá de las b_i

$$I(p_{n-1}) = \Phi(b_1, b_2, \dots, b_{n-1}),$$

y será por consiguiente una función de estas ordenadas. Elegimos ahora las b_i de manera que den, para $I(p_{n-1})$, un valor mínimo. Para determinar estas b_i tendremos un sistema de ecuaciones

$$\frac{\partial}{\partial b_i} I(p_{n-1}) = 0 \quad (i = 1, 2, \dots, n-1).$$

Puesto que toda curva admisible —y en particular la solución exacta del problema— puede ser aproximada por poligonales con un grado de exactitud cualquiera, tanto en su posición sobre el plano como en la dirección de sus tangentes, está claro que la sucesión de poligonales p_{n-1} de este modo construida será, en efecto, una sucesión minimizante. Tomando n suficientemente grande podemos esperar aproximar la solución con el grado deseado de exactitud sobre todo el intervalo $[x_1, x_2]$. Naturalmente, la convergencia debe ser investigada en cada caso.

El siguiente método, que es muy conveniente a efectos de cálculo, es muy utilizado en física y tecnología.

Elegimos una función $\phi_0(x)$ que verifique las condiciones de contorno $\phi_0(x_1) = y_1$ y $\phi_0(x_2) = y_2$, y una sucesión de funciones $\phi_1(x)$, $\phi_2(x)$, $\dots$, que se anulen en los extremos del intervalo $[x_1, x_2]$.

Formamos la combinación lineal

$$s_n(x) = \phi_0(x) + a_1\phi_1(x) + \dots + a_n\phi_n(x).$$

Para valores arbitrarios de los coeficientes numéricos $a_1, a_2, \dots, a_n$, la función $s_n(x)$ será admisible.

Sustituyendo y por $s_n(x)$ en la integral I y haciendo los cálculos necesarios, obtenemos una cierta función de los coeficientes a_i .

Ahora elegimos los a_i tales que esta función tenga el menor valor posible. Los coeficientes deberán ser calculados a partir del sistema

$$\frac{\partial}{\partial a_i} I(s_n) = 0 \quad (i = 1, 2, \dots, n).$$

Resolviendo este sistema, obtenemos, en general, los valores de los coeficientes $a_1, \dots, a_n$ que producen un mínimo para $I(s_n)$, y con ellos construimos una aproximación de la solución

$$\bar{s}_n(x) = \phi_0(x) + \bar{a}_1\phi_1(x) + \dots + \bar{a}_n\phi_n(x).$$

La sucesión de aproximaciones $\bar{s}_n$ ($n = 1, 2, \dots$) construida de esta manera no será minimizante para una elección arbitraria de las funciones ϕ_i . La condición necesaria para ello es que la sucesión de funciones ϕ_i verifique una cierta condición de «complitud» que no definiremos aquí.

Capítulo 9

FUNCIONES DE UNA VARIABLE COMPLEJA

§1. *Números complejos y funciones de una variable compleja*

Números complejos y su significación en álgebra

Los números complejos se introdujeron en matemáticas en conexión con la resolución de ecuaciones algebraicas. La imposibilidad de resolver la ecuación

$$x^2 + 1 = 0 \quad (1)$$

en el campo de los números reales condujo a la introducción de un número convencional, la unidad imaginaria i , definida por la ecuación

$$i^2 = -1. \quad (2)$$

Los números de la forma $a + bi$, donde a y b son reales, se denominaron números *complejos*. Estos se manejaron igual que los reales, sumándolos y multiplicándolos como binomios. Si además utilizamos la ecuación (2), las operaciones básicas de la aritmética, al efectuarlas con números complejos, dan lugar a otros números complejos¹. Definiendo la división como la inversa de la multiplicación,

¹ Los números complejos son conocidos para el lector desde el bachillerato. Véase también el capítulo 4, §3.

resulta que aquélla también está unívocamente determinada, con la única condición de que el denominador sea distinto de cero. De esta manera, la introducción de los números complejos sacó a la luz el hecho interesante —aunque por el momento puramente formal— de que además de los reales existen otros números, los complejos, sobre los cuales se pueden efectuar todas las operaciones aritméticas.

El paso siguiente consiste en la representación geométrica de números complejos. Todo complejo $a + bi$ se puede representar, en el plano Oxy , por un punto de coordenadas (a, b) , o por un vector dirigido desde el origen al punto (a, b) . Esto conduce a un nuevo punto de vista relativo a los números complejos. Los números complejos son pares (a, b) de números reales para los cuales se han establecido definiciones de las operaciones de adición y multiplicación, obedeciendo las mismas leyes que para los números reales. Aquí descubrimos una notable situación: la suma de dos números complejos

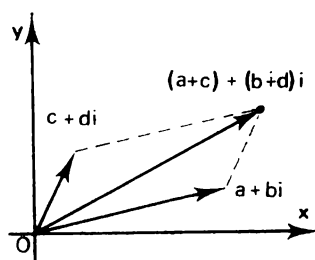


FIG. 1

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

está representada geométricamente por la diagonal del paralelogramo construido a partir de los vectores que representan los sumandos (Fig. 1). De esta forma, los números complejos se suman siguiendo las mismas leyes que las cantidades vectoriales encontradas en mecánica y física: fuerzas, velocidades y aceleraciones. Esta fue una razón más para pensar que los números complejos no son meras generalizaciones formales, sino que pueden ser utilizados para representar magnitudes físicas reales.

Veremos más adelante cómo este punto de vista es muy afortunado en la resolución de distintos problemas de física matemática.

Sin embargo, la introducción de los números complejos se apuntó sus primeros éxitos en el hallazgo de las leyes del álgebra y del análisis. El dominio de los números reales, cerrado con respecto a las operaciones aritméticas, se vio que no era suficientemente amplio para el álgebra. Incluso una ecuación tan sencilla como la (1) no posee una raíz en dicho dominio; para los números complejos tene-

mos en cambio el siguiente hecho notable, llamado teorema fundamental del álgebra: toda ecuación algebraica

$$z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n = 0$$

con coeficientes complejos tiene n raíces complejas².

Este teorema muestra que los complejos forman un sistema de números que, en un sentido bien conocido, es completo con respecto a las operaciones del álgebra. No es de ningún modo trivial que añadiendo al dominio de los números reales una raíz de la ecuación (1) obtengamos los números $a + bi$ en cuyo dominio cualquier ecuación algebraica es resoluble. El teorema fundamental del álgebra mostró que la teoría de polinomios, incluso con coeficientes reales, sólo puede adoptar una forma acabada cuando consideramos los valores del polinomio en todo el plano complejo. El ulterior desarrollo de la teoría de polinomios algebraicos vino a apoyar cada vez más este punto de vista. Las propiedades de los polinomios sólo se ponen de manifiesto al considerarlos como funciones de una variable compleja.

Serie de potencias y funciones de una variable compleja

El desarrollo del análisis puso de manifiesto una serie de hechos que muestran que la introducción de los números complejos era significativa no sólo en la teoría de polinomios, sino también en otra interesantísima clase de funciones, aquellas que son desarrollables en serie de potencias

$$f(x) = a_0 + a_1(x - a) + a_2(x - a)^2 + \cdots \quad (3)$$

Como ya se mencionó en el capítulo 2, el desarrollo del análisis infinitesimal precisó del establecimiento de un punto de vista más exacto para el concepto de función y para las distintas posibilidades de definir funciones en matemáticas. Sin detenernos aquí a discutir estas interesantes cuestiones, recordemos solamente que en los mismos comienzos del desarrollo del análisis se comprobó ya que las funciones que aparecían con más frecuencia podían ser desarrolladas en serie de potencias en el entorno de cada punto del dominio de definición. Esta propiedad es cierta, por ejemplo, para todas las llamadas *funciones elementales*.

La mayoría de los problemas concretos del análisis conducían a funciones que son desarrollables en serie de potencias. Por otra

² Véase el capítulo 4, §3.

parte, existía el deseo de conectar la definición de función «matemática» con una fórmula «matemática», y en este sentido las series de potencias representaban un tipo de fórmula «matemática» muy abarcador. Esta situación condujo incluso a serios intentos de restringir el análisis al estudio de funciones que fueran desarrollables en serie de potencias y que, por tanto, se denominaron *funciones analíticas*. El desarrollo de la ciencia demostró que tal restricción es impropio. Los problemas de física matemática empezaron a extenderse más allá de la clase de funciones analíticas, la cual ni siquiera incluye, por ejemplo, funciones representadas por curvas que poseen angulosidades. Sin embargo, debido a sus notables propiedades y sus numerosas aplicaciones, la clase de funciones analíticas, resultó ser la más importante de todas las clases de funciones estudiadas por los matemáticos.

Como el cálculo de cada término de una serie de potencias requiere solamente operaciones aritméticas, los valores de una función representada por una serie de potencias pueden calcularse también para valores complejos del argumento, al menos para aquellos para los cuales la serie converge. Si ampliamos así la definición de una función de una variable real con el fin de incluir argumentos complejos, hablamos de la «prolongación» de la función en el dominio complejo. De este modo, una función analítica, de la misma forma que un polinomio, puede ser considerada no sólo para valores reales del argumento, sino también para complejos. Además, también podemos considerar series de potencias con coeficientes complejos. Las propiedades de las funciones analíticas, como también las de los polinomios, sólo se ponen completamente de manifiesto cuando se las considera en el dominio complejo. Para ilustrar esto vayamos ahora a un ejemplo.

Consideremos las dos funciones de una variable real

$$e^x \quad \text{y} \quad \frac{1}{1+x^2}.$$

Ambas funciones son finitas, continuas y diferenciables un número arbitrario de veces sobre todo el eje Ox . Se pueden desarrollar en una serie de Taylor, por ejemplo, alrededor del origen $x = 0$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots, \quad (4)$$

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots. \quad (5)$$

La primera de las series así obtenidas converge para todo valor de x , mientras que la segunda converge sólo para $-1 < x < +1$. Considerando la función (5) para valores reales del argumento no se ve por qué su serie de Taylor diverge para $|x| \geq 1$. Pasando al campo complejo podemos poner en claro la situación. Consideremos la serie (5) para valores complejos del argumento

$$1 - z^2 + z^4 - z^6 + \dots \quad (6)$$

La suma de n términos de esta serie

$$s_n = 1 - z^2 + z^4 - z^6 + \dots + (-1)^{n-1} z^{2n-2}$$

se calcula de la misma forma que para valores reales de z :

$$s_n + z^2 s_n = 1 + (-1)^n z^{2n},$$

de donde

$$s_n = \frac{1 + (-1)^n z^{2n}}{1 + z^2}.$$

Esta expresión muestra que para $|z| < 1$

$$\lim_{n \rightarrow \infty} s_n = \frac{1}{1 + z^2},$$

ya que $|z|^{2n} \rightarrow 0$. Así, pues, para complejos z que verifiquen la desigualdad $|z| < 1$ la serie (6) converge y su suma es igual a $1/(1 + z^2)$. Para $|z| \geq 1$ la serie (6) diverge, puesto que en este caso la diferencia $s_n - s_{n-1} = (-1)^{n-1} z^{2n-2}$ no converge a cero.

La desigualdad $|z| < 1$ indica que el punto z está situado a una distancia del origen que es menor que uno. Así, pues, los puntos en los cuales la serie (6) converge forman un círculo en el plano complejo con centro en el origen. Sobre la circunferencia de este círculo están dos puntos i y $-i$, para los cuales la función $1/(1 + z^2)$ se hace infinita; la presencia de estos puntos determina las restricciones sobre el dominio de convergencia de la serie (6).

El dominio de convergencia de una serie de potencias

El dominio de convergencia de la serie de potencias

$$a_0 + a_1(z - a) + a_2(z - a)^2 + \dots + a_n(z - a)^n + \dots \quad (7)$$

en el plano complejo es siempre un círculo con centro en el punto a .

Demostremos esta proposición denominada *teorema de Abel*.

En primer lugar observemos que una serie cuyos términos son los números complejos w_n

$$w_1 + w_2 + \cdots + w_n + \cdots, \quad (8)$$

se puede considerar como dos series, formadas por las partes reales y las partes imaginarias del número $w_n = u_n + iv_n$:

$$u_1 + u_2 + \cdots + u_n + \cdots, \quad (9)$$

$$v_1 + v_2 + \cdots + v_n + \cdots. \quad (10)$$

Una suma parcial s_n de la serie (8) se expresa mediante las sumas parciales σ_n y τ_n de las series (9) y (10)

$$s_n = \sigma_n + i\tau_n,$$

por lo que la convergencia de (8) es equivalente a la convergencia simultánea de (9) y (10), y la suma s de aquélla viene expresada por las sumas σ y τ de (9) y (10)

$$s = \sigma + i\tau.$$

Después de estas observaciones el siguiente lema es trivial:

Si los términos de (8) son menores en valor absoluto que los de una serie geométrica convergente

$$A + Aq + Aq^2 + \cdots + Aq^n + \cdots$$

con A y q positivos, donde $q < 1$, entonces la serie (8) converge.

Pues si $|w_n| < Aq^n$, entonces

$$|u_n| \leq |w_n| < Aq^n,$$

$$|v_n| \leq |w_n| < Aq^n,$$

de modo que (véase capítulo 2, sección 14) las series (9) y (10) convergen y por consiguiente también la (8).

Probemos ahora que, si la serie (7) converge en un punto z_0 , entonces también converge en todos los puntos interiores a un círculo de centro a y que contenga a z_0 en su frontera (Fig. 2). De esta pro-

posición se deduce inmediatamente que el dominio de convergencia de la serie (7)

$$a_0 + a_1(z - a) + \cdots + a_n(z - a)^n + \cdots$$

es, o todo el plano, o el único punto $z = a$, o un círculo de radio finito.

Supongamos que la serie (7) converge en el punto z_0 ; entonces su término general para $z = z_0$ converge a cero cuando $n \rightarrow \infty$, lo cual significa que todos los términos de la serie son interiores a cierto círculo; sea A el radio de dicho círculo, de modo que para cualquier n

$$|a_n(z_0 - a)^n| < A. \quad (11)$$

Tomemos ahora un punto z cualquiera más próximo que z_0 a a , y probemos que en el punto z la serie converge.

Evidentemente,

$$|z - a| < |z_0 - a|,$$

de modo que

$$q = \frac{|z - a|}{|z_0 - a|} < 1. \quad (12)$$

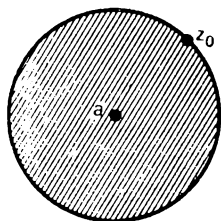


FIG. 2

Estimemos el término general de la serie (7) en el punto z

$$|a_n(z - a)^n| = |a_n(z_0 - a)^n \left(\frac{z - a}{z_0 - a} \right)^n| = |a_n(z_0 - a)^n| \left(\frac{|z - a|}{|z_0 - a|} \right)^n;$$

de las desigualdades (11) y (12) se sigue que

$$|a_n(z - a)^n| < Aq^n;$$

es decir, el término general de la serie (7) en el punto z es menor que el término general de una serie geométrica convergente. Por el lema básico anterior, la serie (7) converge en el punto z .

El círculo dentro del cual una serie de potencias converge, y en cuyo exterior diverge, se denomina *círculo de convergencia*; su radio se denomina *radio de convergencia* de la serie. La frontera del círculo

de convergencia, como puede probarse, siempre pasa por el punto más próximo a a en el que la función deja de tener un comportamiento regular.

La serie de potencias (4) converge en todo el plano complejo; la serie (5), como ya hemos probado, tiene un radio de convergencia igual a uno.

Funciones exponencial y trigonométrica de una variable compleja

Una serie de potencias puede servir para «prolongar» una función de variable real al campo complejo. Por ejemplo, para un valor complejo de z definimos la función e^z mediante la serie de potencias

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots \quad (13)$$

Análogamente las funciones trigonométricas de una variable compleja se introducen mediante las expresiones:

$$\operatorname{sen} z = \frac{z}{1!} - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots, \quad (14)$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots. \quad (15)$$

Estas series convergen en todo el plano.

Es interesante observar la conexión que existe entre las funciones exponencial y trigonométrica cuando nos movemos en el campo complejo.

Si en (13) reemplazamos z por iz , obtenemos

$$e^{iz} = 1 + i \frac{z}{1!} - \frac{z^2}{2!} - i \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots.$$

Agrupados todos los términos sin el factor i por un lado, y los términos con factor i por otro, tenemos

$$e^{iz} = \cos z + i \operatorname{sen} z. \quad (16)$$

Análogamente podemos deducir

$$e^{-iz} = \cos z - i \operatorname{sen} z. \quad (16')$$

Las expresiones (16) y (16') se denominan *fórmulas de Euler*. Despejando $\cos z$ y $\operatorname{sen} z$ en (16) y (16'), obtenemos

$$\begin{aligned} \cos z &= \frac{e^{iz} + e^{-iz}}{2}, \\ \operatorname{sen} z &= \frac{e^{iz} - e^{-iz}}{2i}. \end{aligned} \quad (17)$$

Es muy importante que para valores complejos la sencilla regla de adición de exponentes siga siendo válida

$$e^{z_1} \cdot e^{z_2} = e^{z_1 + z_2}. \quad (18)$$

Puesto que para valores complejos del argumento definimos la función e^z mediante la serie (13), la fórmula (18) debe demostrarse sobre la base de esta definición. A continuación damos la demostración:

$$e^{z_1} \cdot e^{z_2} = \left(1 + \frac{z_1}{1!} + \frac{z_1^2}{2!} + \cdots\right) \cdot \left(1 + \frac{z_2}{1!} + \frac{z_2^2}{2!} + \cdots\right).$$

Efectuaremos la multiplicación de las series término a término. Los términos obtenidos de esta manera se pueden escribir en forma de tabla rectangular

$$\begin{array}{ccccccc} 1 \cdot 1 & + & 1 \cdot \frac{z_2}{1!} & + & 1 \cdot \frac{z_2^2}{2!} & + & 1 \cdot \frac{z_2^3}{3!} + \\ & & \cdots & + & \frac{z_1}{1!} \cdot 1 & + & \frac{z_1}{1!} \cdot \frac{z_2}{1!} + \frac{z_1}{1!} \cdot \frac{z_2^2}{2!} + \frac{z_1}{1!} \cdot \frac{z_2^3}{3!} + \\ & & \cdots & + & \frac{z_1^2}{2!} \cdot 1 & + & \frac{z_1^2}{2!} \cdot \frac{z_2}{1!} + \frac{z_1^2}{2!} \cdot \frac{z_2^2}{2!} + \frac{z_1^2}{2!} \cdot \frac{z_2^3}{3!} + \\ & & \cdots & + & \frac{z_1^3}{3!} \cdot 1 & + & \frac{z_1^3}{3!} \cdot \frac{z_2}{1!} + \frac{z_1^3}{3!} \cdot \frac{z_2^2}{2!} + \frac{z_1^3}{3!} \cdot \frac{z_2^3}{3!} + \cdots \end{array}$$

Agrupemos los términos que tengan la misma suma de potencias de z_1 y z_2 . Es fácil ver que tales términos están sobre las diagonales de nuestra tabla. Obtenemos

$$e^{z_1} \cdot e^{z_2} = 1 + \left(\frac{z_2}{1!} + \frac{z_1}{1!} \right) + \left(\frac{z_2^2}{2!} + \frac{z_1}{1!} \frac{z_2}{1!} + \frac{z_1^2}{2!} \right) + \dots \quad (19)$$

El término general de esta serie será

$$\begin{aligned} \frac{z_2^n}{n!} + \frac{z_2^{n-1}}{(n-1)!} \frac{z_1}{1!} + \frac{z_2^{n-2}}{(n-2)!} \frac{z_1^2}{2!} + \dots + \frac{z_1^n}{n!} = \\ = \frac{1}{n!} \left(z_2^n + \frac{n!}{1!(n-1)!} z_2^{n-1} z_1 + \frac{n!}{2!(n-2)!} z_2^{n-2} z_1^2 + \dots + z_1^n \right). \end{aligned}$$

Aplicando la fórmula del binomio de Newton, obtenemos el término general en la forma

$$\frac{(z_1 + z_2)^n}{n!}.$$

Así, el término general de la serie (19) es idéntico al de la serie de $e^{z_1+z_2}$, lo cual prueba el teorema sobre la regla para la multiplicación (18).

El teorema de multiplicación y la fórmula de Euler nos permiten deducir una expresión para la función e^z en términos de funciones de variables reales en forma finita (sin series). Así, pues, poniendo

$$z = x + iy,$$

obtenemos

$$e^z = e^{x+iy} = e^x \cdot e^{iy},$$

y como

$$e^{iy} = \cos y + i \sin y,$$

hallamos que

$$e^z = e^x(\cos y + i \sin y). \quad (20)$$

La fórmula así deducida es muy conveniente para investigar las propiedades de la función e^z . Señalemos dos de sus propiedades: 1) la función e^z no se anula en ningún punto; ya que, en efecto, $e^x \neq 0$, y las funciones $\cos y$ y $\sin y$ en la fórmula (20) nunca se anulan simultáneamente; 2) la función e^z tiene período $2\pi i$, es decir,

$$e^{z+2\pi i} = e^z.$$

Esto último se deduce del teorema de multiplicación y de la igualdad

$$e^{2\pi i} = \cos 2\pi + i \operatorname{sen} 2\pi = 1.$$

Las fórmulas (17) nos permiten estudiar las funciones $\cos z$ y $\operatorname{sen} z$ en el campo complejo. Dejamos como ejercicio para el lector el probar que en el campo complejo $\cos z$ y $\operatorname{sen} z$ tienen período 2π y que los teoremas sobre el seno y coseno de una suma continúan siendo válidos para ellas.

El concepto general de función de una variable compleja y la diferenciabilidad de funciones

Las series de potencias nos permiten definir funciones analíticas de una variable compleja. Sin embargo, es interesante estudiar las operaciones básicas del análisis para una función arbitraria de una variable compleja y en particular la operación de diferenciación. Pondremos aquí al descubierto hechos muy profundos relacionados con la diferenciación de funciones de una variable compleja. Como veremos, una función que tenga derivada primera en todos los puntos de un entorno de z_0 , tiene necesariamente derivadas de todos los órdenes en z_0 , y además es desarrollable en una serie de potencias centrada en este punto; es decir, es analítica. Así, pues, si consideramos funciones diferenciables de una variable compleja volvemos inmediatamente a la clase de funciones analíticas. Por otra parte, un estudio de la derivada pone al descubierto el comportamiento geométrico de las funciones de una variable compleja y las conexiones de la teoría de estas funciones con problemas de física matemática.

A la vista de lo anterior, diremos, en lo que sigue, que una función es *analítica en el punto* z_0 si tiene derivada en todos los puntos de algún entorno de z_0 .

Siguiendo la definición general de función, diremos que una variable compleja w es una función de la variable compleja z si existe alguna ley que nos permita hallar el valor de w , dado el valor de z .

Cada número complejo $z = x + iy$ está representado por un punto (x, y) en el plano Oxy , y los números $w = u + iv$ estarán representados también por puntos de un plano Ouv , el plano de la función. Entonces, desde el punto de vista geométrico, una función de variable compleja $w = f(z)$ define una ley de correspondencia entre los puntos del plano Oxy del argumento z y puntos del plano Ouv del valor w de la función. En otras palabras, una función de variable compleja determina una transformación del plano del argumento en

el plano de la función. Definir una función de variable compleja significa dar la correspondencia entre los pares de números (x, y) y (u, v) ; definir una función de variable compleja es, por tanto, equivalente a definir dos funciones

$$u = \phi(x, y), \quad v = \psi(x, y),$$

para las cuales, evidentemente,

$$w = u + iv = \phi(x, y) + i\psi(x, y).$$

Por ejemplo, si

$$w = z^2 = (x + iy)^2 = x^2 - y^2 + 2ixy,$$

entonces

$$u = \phi(x, y) = x^2 - y^2, \quad v = \psi(x, y) = 2xy.$$

La derivada de una función de variable compleja se define formalmente de la misma forma que la de una función de una variable real. La derivada es el límite del cociente incremental de la función

$$f'(z) = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}, \quad (21)$$

si este límite existe.

Aun suponiendo que las dos funciones reales u y v que componen $w = f(z)$ tengan derivadas parciales con respecto a x e y , ello no es suficiente para que exista la derivada de la función $f(z)$. El límite del cociente incremental, por regla general, depende de la dirección en que el punto $z' = z + \Delta z$ se aproxima a z (Fig. 3). Para la existencia de la derivada $f'(z)$, es necesario que el límite no dependa de cómo se aproxima z' a z . Consideremos, por ejemplo, el caso en que z' se aproxime a z paralelamente al eje Ox o paralelamente al eje Oy .

En el primer caso

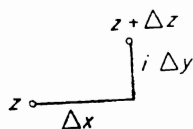


FIG. 3.

$$\Delta z = \Delta x,$$

$$f(z + \Delta z) - f(z) = u(x + \Delta x, y) - u(x, y) + i[v(x + \Delta x, y) - v(x, y)],$$

y el cociente incremental

$$\frac{f(z + \Delta z) - f(z)}{\Delta z} = \frac{u(x + \Delta x, y) - u(x, y)}{\Delta x} + i \frac{v(x + \Delta x, y) - v(x, y)}{\Delta x},$$

para $\Delta x \rightarrow 0$ converge a

$$\frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}. \quad (22)$$

En el segundo caso

$$\Delta z = i\Delta y,$$

y el cociente incremental

$$\frac{f(z + \Delta z) - f(z)}{\Delta z} = -i \frac{u(x, y + \Delta y) - u(x, y)}{\Delta y} + \frac{v(x, y + \Delta y) - v(x, y)}{\Delta y}$$

da, en el límite,

$$\frac{\partial v}{\partial y} - i \frac{\partial u}{\partial y}. \quad (23)$$

Si la función $w = f(z)$ tiene derivada, estas dos expresiones deben ser iguales, y así

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad (24)$$

$$\frac{\partial u}{\partial y} = - \frac{\partial v}{\partial x}.$$

Para que exista la derivada de la función $w = u + iv$ es necesario que se verifiquen estas ecuaciones. Se puede probar que la condición (24) no es solamente necesaria, sino también suficiente (si las funciones u y v tienen una diferencial total). No daremos una demostración de la suficiencia de las condiciones (24), las cuales se denominan *ecuaciones de Cauchy-Riemann*.

Es fácil establecer el hecho de que las reglas usuales para la diferenciación de funciones de una variable real se trasladan sin alteración a funciones de variable compleja. Indudablemente esto es cierto para la derivada de la función z^n y para la derivada de una suma, un producto o un cociente. El procedimiento de la demostración es exactamente el mismo que para las funciones de una variable real, excepto solamente que en lugar de cantidades reales se sobreentiende que son complejas. Esto prueba que todo polinomio en z

$$w = a_0 + a_1 z + \cdots + a_n z^n$$

es una función diferenciable en todo punto. Cualquier función racional igual al cociente de dos polinomios

$$w = \frac{a_0 + a_1 z + \cdots + a_n z^n}{b_0 + b_1 z + \cdots + b_m z^m}$$

es diferenciable en todos los puntos en que el denominador es distinto de cero.

Para establecer la diferenciabilidad de la función $w = e^z$, podemos hacer uso de las condiciones de Cauchy-Riemann. En este caso, apoyándonos en la fórmula (20)

$$u = e^x \cos y, \quad v = e^x \sin y;$$

sustituimos estas funciones en (24) y comprobamos que verifican las ecuaciones de Cauchy-Riemann. La derivada se puede calcular, por ejemplo, mediante la fórmula (22):

$$\frac{dw}{dz} = e^z.$$

Apoyándonos en la fórmula (17) es fácil establecer la diferenciabilidad de las funciones trigonométricas y la validez de las fórmulas conocidas de análisis para hallar los valores de sus derivadas.

La función $\text{Ln } z$

No haremos aquí un estudio de todas las funciones elementales de variable compleja. Sin embargo, es importante para nuestros propósitos llegar a familiarizarnos con las propiedades de la función $\text{Ln } z$. Como en el caso del campo real, ponemos

$$w = \text{Ln } z,$$

si

$$z = e^w.$$

Para analizar la función $\text{Ln } z$, escribimos el número z en forma trigonométrica

$$z = r(\cos \phi + i \sin \phi).$$

Aplicando el teorema de multiplicación a e^w , obtenemos

$$z = e^w = e^{u+iv} = e^u e^{iv} = e^u (\cos v + i \sin v).$$

Igualando las dos expresiones de z , tenemos

$$e^u = r, \quad (a)$$

$$\cos v + i \operatorname{sen} v = \cos \phi + i \operatorname{sen} \phi. \quad (\beta)$$

Como u y r son números reales, de la fórmula (a) deducimos

$$u = \ln r,$$

donde $\ln r$ es el valor usual del logaritmo natural de un número real. La ecuación (β) puede ser satisfecha solamente si

$$\cos v = \cos \phi, \quad \operatorname{sen} v = \operatorname{sen} \phi,$$

y en este caso v y ϕ deben diferir en un múltiplo de 2π

$$v = \phi + 2\pi n,$$

donde para cualquier entero n la ecuación (β) será verificada. Apoyándonos en las expresiones deducidas para u y v

$$\operatorname{Ln} z = \ln r + i(\phi + 2\pi n). \quad (25)$$

La fórmula (25) define la función $\operatorname{Ln} z$ para todos los valores del número complejo z que sean distintos de cero. Da la definición del logaritmo no sólo para números positivos, sino también para negativos y complejos.

La expresión deducida para la función $\operatorname{Ln} z$ contiene un entero arbitrario n . Esto significa que $\operatorname{Ln} z$ es una función multiforme. Para cualquier valor de n obtenemos uno de los posibles valores de la función $\operatorname{Ln} z$. Si fijamos el valor de n , obtenemos uno de los posibles valores de esta función.

Sin embargo, los diferentes valores de $\operatorname{Ln} z$, como se puede demostrar, están orgánicamente relacionados unos con otros. En efecto, fijemos, por ejemplo, el valor $n = 0$ en el punto z_0 y supongamos que z se mueve continuamente a lo largo de una curva cerrada C , la cual rodea el origen y vuelve al punto z_0 (Fig. 4). Durante el movi-

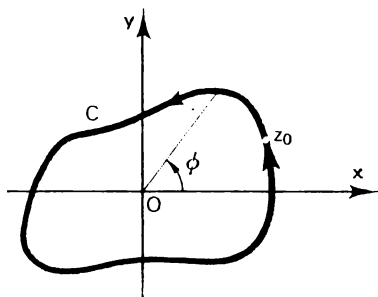


FIG. 4

miento de z , el ángulo ϕ crecerá continuamente, y cuando z haya recorrido todo el contorno cerrado, ϕ habrá aumentado en 2π . De esta manera, fijando el valor del logaritmo en z_0

$$(\operatorname{Ln} z)_0 = \ln r_0 + i\phi_0,$$

y variando este valor continuamente mientras se mueve z a lo largo de la curva cerrada que rodea al origen, volvemos al punto z_0 con un valor distinto de la función

$$(\operatorname{Ln} z)_0 = \ln r_0 + i(\phi_0 + 2\pi).$$

Esta situación nos prueba que podemos pasar de un valor de $\operatorname{Ln} z$ a otro de una forma continua. Para ello sólo es necesario que el punto se mueva continuamente alrededor del origen un número suficiente de veces. El punto $z = 0$ se denomina *punto de ramificación* de la función $\operatorname{Ln} z$.

Si deseamos restringir nuestra consideración a un único valor de la función $\operatorname{Ln} z$, debemos evitar que el punto z describa una curva cerrada que rodee el punto $z = 0$. Esto puede hacerse trazando una curva continua desde el origen al infinito y evitando que el punto z atraviese esta curva, la cual se denomina un *corte*. Si z varía sobre el plano cortado, entonces ya no cambia continuamente de un valor de $\operatorname{Ln} z$ a otro y, por tanto, partiendo de un valor específico del logaritmo en un punto z_0 , obtenemos en cada punto solamente un valor del logaritmo. Los valores de la función $\operatorname{Ln} z$ seleccionados de esta manera constituyen una rama uniforme de la misma.

Por ejemplo, si el corte se halla a lo largo de la parte negativa del eje Ox , obtenemos una rama uniforme de $\operatorname{Ln} z$ restringiendo el argumento a los límites

$$(2k - 1)\pi < \phi < (2k + 1)\pi,$$

donde k es un entero arbitrario.

Considerando una rama uniforme del logaritmo, podemos estudiar su diferenciabilidad. Poniendo

$$r = \sqrt{x^2 + y^2}, \quad \phi = \arctg \frac{y}{x},$$

es fácil probar que $\text{Ln } z$ verifica las condiciones de Cauchy-Riemann, y su derivada —calculada por ejemplo mediante la fórmula (22)— será igual a

$$\frac{d \text{Ln } z}{dz} = \frac{1}{z}.$$

Hagamos hincapié en que la derivada de $\text{Ln } z$ es también una función uniforme.

§2. *La conexión entre las funciones de una variable compleja y los problemas de la física matemática*

Conexión con problemas de hidrodinámica

Las condiciones de Cauchy-Riemann establecen una relación entre los problemas de física matemática y la teoría de funciones de una variable compleja. Ilustremos esto con algunos problemas de hidrodinámica.

Entre todos los posibles movimientos de un fluido juegan un importante papel los llamados movimientos *permanentes*, es decir, aquellos para los cuales la distribución de velocidades del fluido en el espacio no cambia con el tiempo. Por ejemplo, un observador que, situado sobre un puente, contempla la corriente del río en torno a uno de los pilares verá un flujo permanente. A veces, una corriente es permanente para un observador en movimiento sobre algún vehículo. En el caso de un vapor que navegue por aguas agitadas, la corriente le parecerá no permanente a un observador situado en la orilla, pero permanente a un pasajero del barco. Para un pasajero de un avión que vuele con velocidad constante, la corriente de aire, aunque perturbada por el avión, le seguirá pareciendo permanente.

En un movimiento permanente, el vector velocidad V de la partícula de fluido que pasa por un punto dado del espacio, no cambia con el tiempo. Si el movimiento es permanente para un observador móvil, entonces el vector velocidad no cambia con el tiempo en puntos que tengan coordenadas constantes en un sistema de coordenadas que se mueva con el observador.

Entre los movimientos de un fluido se ha atribuido gran importancia a la clase de *movimientos laminares*. Son corrientes para las cuales la velocidad de las partículas es en todos los puntos paralela

a algún plano, y la distribución de velocidades es idéntica en todos los planos paralelos al dado.

Si imaginamos una masa de fluido infinitamente extensa que pase alrededor de un cuerpo cilíndrico en una dirección perpendicular a una generatriz, la distribución de velocidades será la misma sobre todos los planos perpendiculares a la generatriz, por lo que el flujo será laminar. En muchos casos el movimiento de un fluido es aproximadamente laminar. Por ejemplo, si consideramos el flujo de aire en un plano perpendicular al ala de un avión, el movimiento del aire se puede considerar aproximadamente como laminar, con tal de que el plano en cuestión no esté demasiado próximo al fuselaje ni al extremo del ala.

Veremos cómo la teoría de funciones de una variable compleja puede ser aplicada al estudio del flujo laminar permanente.

Aquí supondremos que el líquido es incompresible, es decir, que su densidad no cambia al variar la presión. Esta hipótesis es válida, por ejemplo, para el agua, pero se puede probar que, a efectos del estudio del flujo, incluso el aire se puede considerar incompresible, siempre que la velocidad del movimiento no sea muy grande. La hipótesis de la incompresibilidad del aire no produce una notable tergiversación si las velocidades del movimiento no exceden del orden de 0,6 a 0,8 veces la velocidad del sonido (330 m/seg).

El flujo de un líquido se caracteriza por la distribución de las velocidades de sus partículas. Si el flujo es laminar, entonces basta determinar las velocidades de las partículas en uno de los planos paralelamente a los cuales tiene lugar el movimiento.

Denotaremos por $V(x, y, t)$ el vector velocidad de la partícula que pasa por el punto de coordenadas x, y en el instante t . En el caso de régimen permanente, V no depende del tiempo. El vector V estará dado por sus proyecciones u y v sobre los ejes de coordenadas. Consideramos la trayectoria de las partículas del fluido. En el caso de régimen permanente, no hay ninguna variación temporal en las velocidades de las sucesivas partículas que provienen de un punto dado del espacio. Si conocemos el campo de velocidades, es decir, si conocemos las componentes de la velocidad como funciones de x, y , entonces las trayectorias de las partículas se pueden determinar utilizando el hecho de que la velocidad de cada una de ellas es tangente a la trayectoria. Así, pues, tenemos que

$$\frac{dy}{dx} = \frac{v(x, y)}{u(x, y)}.$$

La expresión así obtenida es la ecuación diferencial para las trayectorias. La trayectoria de una partícula en un movimiento permanente se denomina *línea de corriente*. Por cada punto del plano pasa exactamente una línea de corriente.

Juega aquí un papel importante la llamada función de corriente. Para una línea de corriente fija C_0 consideremos el canal imaginario limitado por las cuatro paredes siguientes: una es la superficie cilíndrica (con generatrices perpendiculares al plano de la corriente) que pasa por la línea de corriente C_0 ; la segunda pared es la misma superficie cilíndrica para una línea de corriente próxima C_1 ; la tercera es el plano de la corriente, y la cuarta es un plano paralelo al anterior a la distancia unidad (Fig. 5). Si consideramos dos secciones transversales arbitrarias de nuestro canal, denotadas por γ_1 y γ_2 , entonces la cantidad de fluido que pasa a través de las secciones γ_1 y γ_2 en la unidad de tiempo será la misma, como se deduce del hecho de que la cantidad de fluido interior a la parte de canal delimitada por C_1 , C_0 y γ_1 , γ_2 no puede cambiar, a causa de la densidad constante, ya que las paredes laterales del canal C_0 y C_1 están formadas por líneas de corriente, por lo que no pasa fluido alguno a través de ellas. En consecuencia, la cantidad de fluido que sale en la unidad de tiempo a través de γ_1 es igual a la que entra a través de γ_2 .

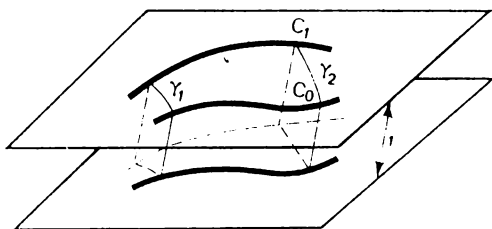


FIG. 5

Se llama *función de corriente* a la función $\psi(x, y)$ que tiene un valor constante sobre la línea de corriente C_1 igual a la cantidad de fluido que pasa a través de una sección transversal del canal construida sobre las curvas C_0 y C_1 , en la unidad de tiempo.

La función de corriente está determinada salvo una constante arbitraria, que depende de la elección de la línea de corriente inicial C_0 . Si conocemos la función de corriente, entonces las ecuaciones para las líneas de corriente son evidentemente

$$\psi(x, y) = \text{const.}$$

Ahora deseamos expresar las componentes de la velocidad de la corriente en un punto dado $M(x, y)$ en función de las derivadas de la función de corriente. Para ello consideremos el canal formado por la línea de corriente C que pasa por el punto $M(x, y)$ y otra línea próxima, C' , que pasa por un punto cercano $M'(x, y + \Delta y)$, junto con los planos paralelos al de corriente que disten una unidad de él. Calculemos la cantidad de fluido q que pasa a través de la sección MM' del canal durante el tiempo dt .

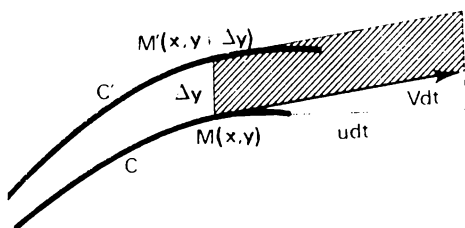


FIG. 6

durante el tiempo dt .

Por una parte, de la definición de función de corriente tenemos:

$$q = (\psi' - \psi) dt.$$

Por otra, q es igual (Fig. 6) al volumen del sólido formado trazando el vector

$V dt$ desde cada punto de la sección MM' . Si esta sección es pequeña, podemos suponer que V es constante sobre el total de MM' y es igual al valor de V en el punto M . El área de la base del paralelepípedo así construido es $\Delta y \times 1$ (en la figura 6 no se muestra el espesor unidad), y la altura es la proyección del vector $V dt$ sobre el eje Ox , es decir, $u dt$, por lo que

$$q \approx u \Delta y dt$$

y por consiguiente

$$u \Delta y \approx \Delta \psi.$$

Dividiendo esta ecuación por Δy y pasando al límite, obtenemos

$$u = \frac{\partial \psi}{\partial y}. \quad (26)$$

Un razonamiento similar da para la segunda componente

$$v = - \frac{\partial \psi}{\partial x}. \quad (26')$$

Para determinar el campo de los vectores velocidad introducimos, además de la función de corriente, otra función que surge al consi-

derar la rotación de pequeñas partículas de líquido. Imaginemos que una cierta partícula del fluido se solidificase; entonces esta partícula tendría en general un movimiento rotatorio. Sin embargo, si el movimiento del fluido parte del reposo y no existen fricciones internas entre las partículas, entonces se puede probar que éstas no comienzan a rotar. Los movimientos de un fluido en el que no existe rotación de este tipo se denominan irrotacionales; juegan un papel fundamental en el movimiento de sólidos en un fluido. En la teoría de la hidromecánica se prueba que para flujo irrotacional existe una segunda función $\phi(x, y)$ tal que las componentes de la velocidad se expresan mediante las fórmulas

$$u = \frac{\partial \phi}{\partial x}, \quad v = \frac{\partial \phi}{\partial y}; \quad (27)$$

la función ϕ se denomina *potencial de velocidad* del flujo. Más adelante consideraremos movimientos con potencial de velocidad.

Comparando las fórmulas para las componentes de la velocidad a partir de la función de corriente y del potencial de velocidad, obtendremos el siguiente resultado notable.

El potencial de velocidad $\phi(x, y)$ y la función de corriente $\psi(x, y)$ para el flujo de un fluido incompresible verifican las ecuaciones de Cauchy-Riemann:

$$\begin{aligned} \frac{\partial \phi}{\partial x} &= \frac{\partial \psi}{\partial y}, \\ \frac{\partial \phi}{\partial y} &= -\frac{\partial \psi}{\partial x}. \end{aligned} \quad (28)$$

En otras palabras, la función de variable compleja

$$w = \phi(x, y) + i\psi(x, y)$$

es una función diferenciable de variable compleja. Recíprocamente, si tomamos una función cualquiera de variable compleja que sea diferenciable, sus partes real e imaginaria verifican las condiciones de Cauchy-Riemann y se pueden considerar como el potencial de velocidad y la función de corriente del flujo de un fluido incompresible. La función w se denomina *función característica del flujo*.

Consideremos ahora el significado de la derivada de w . Utilizando, por ejemplo, la fórmula (22), tenemos

$$\frac{dw}{dz} = \frac{\partial \phi}{\partial x} + i \frac{\partial \psi}{\partial x}.$$

De (27) y (26') tenemos que:

$$\frac{dw}{dz} = u - iv$$

o, tomando el complejo conjugado

$$u + iv = \left(\overline{\frac{dw}{dz}} \right),$$

donde la barra sobre dw/dz denota el complejo conjugado.

En consecuencia, el vector velocidad del flujo es igual al conjugado del valor de la derivada de la función característica del flujo.

Ejemplos de flujo laminar de un fluido

Consideremos varios ejemplos. Sea

$$w = Az, \tag{30}$$

donde A es un número complejo. De (29) se deduce que

$$u + iv = \bar{A}.$$

Así, pues, la función lineal (30) define el flujo de un fluido con vector velocidad constante. Si ponemos

$$A = u_0 - iv_0,$$

entonces, descomponiendo w en sus partes real e imaginaria, tenemos

$$\phi(x, y) = u_0 x + v_0 y,$$

$$\psi(x, y) = u_0 y - v_0 x,$$

por lo que las líneas de corriente serán rectas paralelas al vector velocidad (Fig. 7).

Como segundo ejemplo consideremos la función

$$w = Az^2,$$

donde la constante A es real. Para representar la gráfica del flujo, determinamos primero las líneas de corriente. En este caso

$$\psi(x, y) = 2Axy,$$

y las ecuaciones de las líneas de corriente son

$$xy = \text{const.}$$

Se trata de hipérbolas que tienen los ejes de coordenadas como asíntotas (Fig. 8). Las flechas indican la dirección del movimiento de las partículas a lo largo de las líneas de corriente para $A > 0$. Los ejes Ox y Oy son también líneas de corriente.



FIG. 7

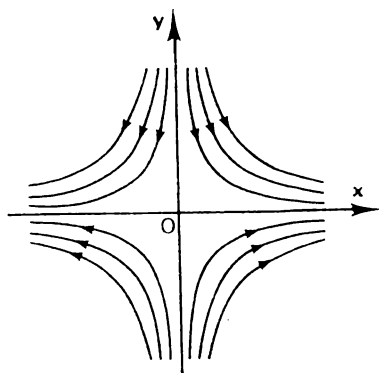


FIG. 8

Si el rozamiento en el líquido es muy pequeño, no perturbaremos el resto del flujo si reemplazamos una línea de corriente por una pared rígida, ya que el fluido resbalará a lo largo de ella. Utilizando este principio para construir paredes a lo largo de los ejes coordenados positivos (en la figura 8 están representadas por líneas gruesas), tenemos un diagrama de cómo un fluido discurre irrotacionalmente, en este caso alrededor de una esquina.

Un ejemplo importante de flujo está dado por la función

$$w = a \left(z + \frac{R^2}{z} \right), \quad (31)$$

donde a y R son cantidades reales positivas.

La función de corriente será

$$\psi = a \left(y - \frac{R^2 y}{x^2 + y^2} \right),$$

y por tanto la ecuación de las líneas de corriente es

$$y - \frac{R^2 y}{x^2 + y^2} = \text{const.}$$

En particular, tomando la constante igual a cero, tenemos $y = 0$ o $x^2 + y^2 = R^2$; así, pues, un círculo de radio R es una línea de corriente. Si sustituimos el interior de esta línea de corriente por un cuerpo sólido, obtenemos el flujo alrededor de un cilindro circular. En la figura 9 se muestra un diagrama de las líneas de corriente de este flujo. La velocidad de la corriente se puede determinar a partir de la fórmula (29) por

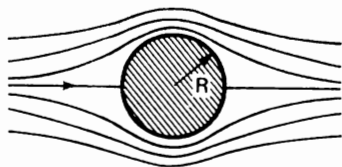


FIG. 9

$$u + iv = a \left(1 - \frac{R^2}{z^2} \right).$$

A gran distancia del cilindro encontramos que

$$\lim_{z \rightarrow \infty} (u + iv) = a;$$

es decir, lejos del cilindro la velocidad tiende a un valor constante y por tanto el flujo tiende a ser uniforme. Por consiguiente, la fórmula (29) define el flujo que se origina por el paso, en torno a un cilindro circular, de un fluido que se mueve uniformemente a cierta distancia de él.

Las ideas fundamentales de la teoría de un ala de avión; teorema de Sukovski

La aplicación de la teoría de funciones de variable compleja al estudio de movimientos laminares de un fluido ha sido la fuente de varios descubrimientos notables en aerodinámica, debidos a Sukovski

y Chaplygin. El estudio de las líneas de corriente de cuerpos les llevó a encontrar la ley de formación de la fuerza de sustentación sobre el ala de un avión. Para exponer las ideas que condujeron a este descubrimiento, necesitamos considerar otro ejemplo concreto de flujo. Sea la función característica

$$w = \frac{\Gamma}{2\pi i} \ln z,$$

donde Γ es una constante real. Si bien w es una función multiforme, su derivada

$$\frac{dw}{dz} = \frac{\Gamma}{2\pi i} \frac{1}{z} \quad (32)$$

es uniforme, por lo que nuestra función determina unívocamente el campo de velocidades de cierta corriente de fluido. Si ponemos $z = re^{i\theta}$, el potencial de velocidad y la función de corriente se pueden calcular aplicando (25), obteniendo

$$\phi = \frac{\Gamma}{2\pi} \theta, \quad \psi = -\frac{\Gamma}{2\pi} \ln r.$$

La segunda de estas fórmulas muestra que las líneas de corriente son los círculos $r = \text{const.}$ (Fig. 10).

La velocidad de la corriente está determinada por la fórmula (29)

$$u + iv = -\frac{\Gamma}{2\pi i} \frac{1}{\bar{z}}.$$

En particular, se deduce que el valor del vector velocidad será

$$V = |u + iv| = \frac{|\Gamma|}{2\pi} \frac{1}{r},$$

es decir, la velocidad es constante sobre cada línea de corriente. Un estudio más detallado muestra que el flujo va en sentido contrario a las agujas del reloj para $\Gamma > 0$, y en el sentido de las mismas, para $\Gamma < 0$.

Si reemplazamos una de las líneas de corriente por un contorno rígido, obtenemos el movimiento circular de un fluido alrededor de un cilindro. Un movimiento de este tipo se denomina *circulatorio*.

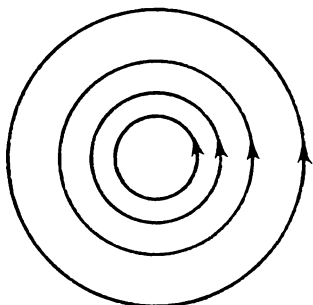


FIG. 10

Sin embargo, el potencial de nuestro movimiento no es una función uniforme. A cada vuelta completa alrededor del cilindro el potencial varía en una cantidad Γ . Esta variación del potencial se denomina *circulación del flujo*.

Si a la función característica de una corriente alrededor de un cilindro (31) le sumamos la de un flujo circulatorio (en el sentido de las agujas del reloj), obtenemos una nueva función característica

$$w = a \left(z + \frac{R^2}{z} \right) - \frac{\Gamma}{2\pi i} \ln z. \quad (33)$$

Esta función también representa el flujo en torno a un cilindro de radio R . En efecto, la función de corriente será constante sobre una circunferencia de radio R , ya que allí los coeficientes de las partes ima-

ginarias de ambos términos son constantes. La velocidad de la corriente, definida por la función (33), converge otra vez a a cuando $z \rightarrow \infty$. Esto prueba que la función característica (33) define, para cualquier valor de Γ , las líneas de corriente de un flujo de traslación alrededor de un cilindro. La figura 11 ilustra el carácter del flujo para $\Gamma > 0$.

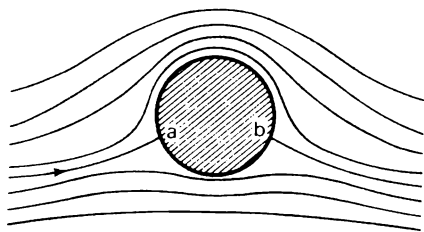


FIG. 11

Este flujo no será simétrico, ya que los puntos de estancamiento a y b , donde la corriente toma contacto y abandona al cilindro, están desplazados hacia abajo. El potencial del flujo en consideración será una función multiforme: al describir una vuelta alrededor del cilindro variará en una cantidad igual a Γ .

Debido a la simetría, el flujo en torno a un cilindro será, normalmente, de la forma determinada por las funciones (32), pero para cuerpos asimétricos el flujo que generalmente se origina tiene un potencial multiforme. Más adelante discutiremos el significado físico de este hecho. Los métodos de la teoría de funciones de variable compleja nos permiten determinar los posibles flujos en torno a cuerpos de forma arbitraria. Estos métodos se discutirán en la siguiente sección. Con su ayuda podemos utilizar el flujo en torno a un cilindro para obtener el flujo —con potencial uniforme o multiforme— en torno a cualquier cuerpo.

Al estudiar las líneas de corriente del ala de avión, estamos tratando en esencia con un cuerpo que posee un borde agudo en la parte posterior. El perfil del ala de avión siempre se estrecha hacia la parte de atrás. Si para un perfil de esta especie construimos un flujo con un potencial uniforme, entonces el punto de estancamiento en que la corriente abandona el ala resulta que no está en el borde (Fig. 12 a). Pero sucede que un flujo así es físicamente imposible (pues en el borde posterior se produciría una velocidad infinita, con

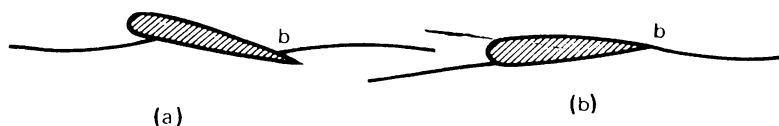


FIG 12

el consiguiente enrarecimiento infinito del fluido en ese punto). El flujo para el que el punto b cae sobre el borde del ala (Fig. 12 b) es el único flujo posible, y éste, por regla general, tiene un potencial multiforme, es decir, es un flujo circulatorio. Su circulación Γ se define de nuevo como la variación del potencial para una vuelta completa alrededor del ala.

La posibilidad física de un flujo alrededor del perfil de un ala con una corriente que abandone dicha ala por el borde trasero se denomina *postulado de Chaplygin*.

El notable descubrimiento de Sukovski es que la existencia de circulación en el flujo origina una fuerza de sustentación sobre el ala, en una dirección perpendicular a la velocidad del flujo e igual en magnitud a la cantidad.

$$\rho a \Gamma,$$

donde ρ es la densidad del medio y Γ la circulación (Fig. 13).

Este teorema de Sukovski acerca de la fuerza de sustentación sobre un ala es fundamental para toda la aerodinámica de nuestros días. No daremos la demostración aquí, únicamente apuntaremos que las demostraciones usuales están basadas en la teoría de integración de funciones de variable compleja.

Los resultados fundamentales de la aerodinámica, tal como los establecieron Sukovski y Chaplygin, han sido extensamente desarrollados por el trabajo de los científicos soviéticos.

Aplicaciones a otros problemas de la física matemática

La teoría de funciones de una variable compleja ha encontrado amplia aplicación no solamente en la teoría del ala, sino en muchos otros problemas de hidrodinámica.

Sin embargo, el dominio de aplicación de la teoría de funciones tampoco está restringido a la hidrodinámica; es mucho más amplio que esto e incluye otros muchos problemas de física matemática.

Para ilustrarlo, volvamos a las condiciones de Cauchy-Riemann

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y},$$

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x},$$



FIG. 13

y deduzcamos de ellas una ecuación que verifica la

parte real de una función analítica de variable compleja. Si se deriva con respecto a x la primera de estas ecuaciones y con respecto a y la segunda, sumando obtenemos

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Esta ecuación (que ya hemos encontrado en el capítulo 6) se conoce como la *ecuación de Laplace*, que aparece en numerosos problemas de física y mecánica. Por ejemplo, si el calor en un cuerpo está en equilibrio, la temperatura verifica la ecuación de Laplace. El estudio de campos magnéticos o electrostáticos está relacionado con esta ecuación. En la investigación de la filtración de un líquido a través de un medio poroso también llegamos a la ecuación de Laplace. En todos estos problemas en que aparece la ecuación de Laplace los métodos de la teoría de funciones han encontrado amplia aplicación.

Aparte de la ecuación de Laplace, también las ecuaciones más generales de la física matemática se pueden relacionar con la teoría de funciones de una variable compleja. Uno de los ejemplos más notables nos lo proporcionan los problemas planares en la teoría

de la elasticidad. Los fundamentos de la aplicación de funciones de variable compleja a este campo fueron colocados por los científicos soviéticos G. B. Kolosov y N. I. Mushelishvili.

§3. *La conexión entre las funciones de variable compleja y la geometría*

Propiedades geométricas de las funciones diferenciables

Como en el caso de funciones de una variable real, en la teoría de funciones analíticas de variable compleja juega un gran papel la interpretación geométrica de estas funciones. En términos generales, las propiedades geométricas de las funciones de variable compleja no sólo proporcionan un medio natural de visualizar las propiedades analíticas de las funciones, sino que también han dado lugar a una clase especial de problemas. La clase de problemas relacionados con las propiedades geométricas de las funciones ha sido denominada *teoría geométrica de funciones*. Como dijimos anteriormente, desde el punto de vista geométrico una función de variable compleja $w = f(z)$ es una transformación del plano z en el plano w . Esta transformación puede también definirse por dos funciones de dos variables reales

$$u = u(x, y),$$

$$v = v(x, y).$$

Si deseamos estudiar el carácter de la transformación en un entorno muy pequeño de un punto, podemos desarrollar estas funciones en serie de Taylor y restringirnos a los términos principales del desarrollo

$$u - u_0 = \left(\frac{\partial u}{\partial x} \right)_0 (x - x_0) + \left(\frac{\partial u}{\partial y} \right)_0 (y - y_0) + \cdots,$$

$$v - v_0 = \left(\frac{\partial v}{\partial x} \right)_0 (x - x_0) + \left(\frac{\partial v}{\partial y} \right)_0 (y - y_0) + \cdots,$$

donde las derivadas se toman en el punto (x_0, y_0) . Por tanto, en las proximidades de un punto cualquier transformación se puede considerar como una transformación afín³

$$u - u_0 = a(x - x_0) + b(y - y_0),$$

$$v - v_0 = c(x - x_0) + d(y - y_0),$$

³ Véase el capítulo 3, §11.

donde

$$a = \left(\frac{\partial u}{\partial x} \right)_0, \quad b = \left(\frac{\partial u}{\partial y} \right)_0,$$

$$c = \left(\frac{\partial v}{\partial x} \right)_0, \quad d = \left(\frac{\partial v}{\partial y} \right)_0.$$

Consideremos las propiedades de la transformación efectuada por la función analítica cerca del punto $z = x + iy$. Sea C una curva que parta del punto z ; sobre el plano w los puntos correspondientes describen la curva Γ , que parte del punto w . Si z' es un punto próximo y w' el transformado, entonces para $z' \rightarrow z$ tendremos $w' \rightarrow w$ y

$$\frac{w' - w}{z' - z} \rightarrow f'(z). \quad (34)$$

En particular, se deduce que

$$\frac{|w' - w|}{|z' - z|} \rightarrow |f'(z)|. \quad (35)$$

Este hecho se puede formular de la manera siguiente.

El límite de la razón de las longitudes de dos cuerdas correspondientes, en el plano w y en el plano z , en el punto z es la misma para todas las curvas que partan de este punto, o dicho de otra manera, la razón de elementos lineales sobre el plano w y sobre el plano z en un punto dado no dependen de la curva que parte de z .

La cantidad $|f'(z)|$, que caracteriza el aumento de elementos lineales en el punto z , se denomina *coeficiente de dilatación* en el punto z .

Supongamos ahora que en algún punto z la derivada $f'(z) \neq 0$, por lo que $f'(z)$ tiene un argumento unívocamente determinado⁴. Calculemos este argumento utilizando (34)

$$\arg \frac{w' - w}{z' - z} = \arg(w' - w) - \arg(z' - z);$$

ahora bien, $\arg(w' - w)$ es el ángulo β' entre la cuerda ww' y el eje real, y $\arg(z' - z)$ es el ángulo α' entre la cuerda zz' y el eje real.

⁴ Véase el capítulo 4, §3.

Si denotamos por α y β los correspondientes ángulos para las tangentes a las curvas C y Γ en los puntos z y w (Fig. 14), entonces, para $z' \rightarrow z$,

$$\alpha' \rightarrow \alpha, \quad \beta' \rightarrow \beta,$$

por lo que en el límite obtenemos

$$\arg f'(z) = \beta - \alpha. \quad (36)$$

Esta ecuación muestra que $\arg f'(z)$ es igual al ángulo ϕ en que hay que rotar la dirección de la tangente a la curva C en el punto z para obtener la dirección de la tangente a la curva Γ en el punto w . De

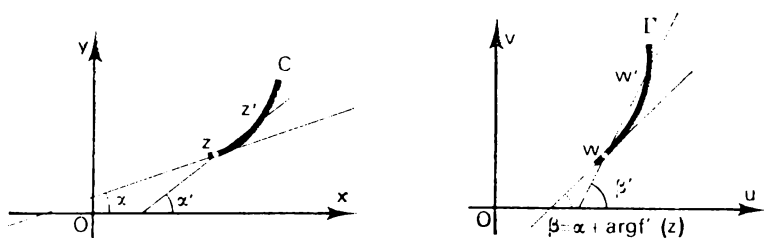


FIG. 14

ahí que $\arg f'(z)$ se denomine *rotación de la transformación* en el punto z .

A partir de la ecuación (36) el lector puede deducir fácilmente las siguientes proposiciones.

Al pasar del plano z al plano w , las tangentes a todas las curvas que parten de un punto dado giran el mismo ángulo.

Si C_1 y C_2 son dos curvas que parten del punto z , y Γ_1 y Γ_2 son las curvas correspondientes desde el punto w , entonces el ángulo entre Γ_1 y Γ_2 en w es igual al ángulo entre C_1 y C_2 en el punto z .

De este modo, en la transformación efectuada por una función analítica (en puntos donde $f'(z) \neq 0$) todos los elementos lineales varían en la misma proporción, y los ángulos entre direcciones correspondientes no cambian.

Las transformaciones con estas propiedades se denominan transformaciones *conformes*.

De las propiedades geométricas que hemos demostrado para transformaciones cerca de un punto en el que $f'(z) \neq 0$, es natural esperar que en un entorno pequeño de z_0 la transformación sea «uno a uno»;

es decir, no solamente corresponderá cada punto z a un solo punto w , sino también recíprocamente, cada punto w corresponderá a un solo punto z . Esta proposición se puede probar rigurosamente.

Para hacer ver con más claridad cómo se distinguen las transformaciones conformes de otros varios tipos de transformaciones, es útil pensar en una transformación arbitraria en un entorno pequeño de un punto. Si consideramos los primeros términos del desarrollo de Taylor de las funciones u y v que efectúan la transformación, obtenemos

$$\begin{aligned} u - u_0 &= \left(\frac{\partial u}{\partial x} \right)_0 (x - x_0) + \left(\frac{\partial u}{\partial y} \right)_0 (y - y_0) + \cdots, \\ v - v_0 &= \left(\frac{\partial v}{\partial x} \right)_0 (x - x_0) + \left(\frac{\partial v}{\partial y} \right)_0 (y - y_0) + \cdots. \end{aligned}$$

Si en un entorno pequeño del punto (x_0, y_0) ignoramos los términos de orden superior, entonces nuestra transformación se comportará como una transformación afín. Esta tiene inversa si su determinante no se anula

$$\Delta = \begin{vmatrix} \left(\frac{\partial u}{\partial x} \right)_0 & \left(\frac{\partial v}{\partial x} \right)_0 \\ \left(\frac{\partial u}{\partial y} \right)_0 & \left(\frac{\partial v}{\partial y} \right)_0 \end{vmatrix} \neq 0.$$

Si $\Delta = 0$, entonces para describir el comportamiento de la transformación en las proximidades del punto (x_0, y_0) debemos considerar términos de orden superior⁵.

En el caso de que $u + iv$ sea una función analítica, podemos expresar las derivadas con respecto a y en función de las derivadas con respecto a x utilizando las condiciones de Cauchy-Riemann, y obtenemos

$$\Delta = \left(\frac{\partial u}{\partial x} \right)_0^2 + \left(\frac{\partial v}{\partial x} \right)_0^2 = \left| \left(\frac{\partial u}{\partial x} \right)_0 + i \left(\frac{\partial v}{\partial x} \right)_0 \right|^2 = |f'(z_0)|^2,$$

es decir, la transformación tiene inversa cuando $f'(z_0) \neq 0$. Si ponemos $f'(z_0) = r(\cos \phi + i \sin \phi)$, entonces

$$\begin{aligned} \left(\frac{\partial u}{\partial x} \right)_0 &= \left(\frac{\partial v}{\partial y} \right)_0 r \cos \phi, \\ \left(\frac{\partial u}{\partial y} \right)_0 &= - \left(\frac{\partial v}{\partial x} \right)_0 = -r \sin \phi, \end{aligned}$$

⁵ Cuando $\Delta = 0$, la transformación no se llama afín. Para transformaciones afines véase también el capítulo 3, §11.

y la transformación, en las proximidades del punto (x_0, y_0) , tendrá la forma

$$u - u_0 = r[(x - x_0) \cos \phi - (y - y_0) \sin \phi] + \dots,$$

$$v - v_0 = r[(x - x_0) \sin \phi + (y - y_0) \cos \phi] + \dots.$$

Estas fórmulas muestran que en el caso de una función analítica $w = u + iv$, la transformación en las proximidades del punto (x_0, y_0) consiste en una rotación de amplitud ϕ y una dilatación de coeficiente r . En efecto, las expresiones interiores a los corchetes son las fórmulas (bien conocidas por geometría analítica) de la rotación en el plano de amplitud ϕ , y la multiplicación por r da la dilatación.

Para formarse una idea de las posibilidades cuando $f'(z) = 0$ es útil considerar la función

$$w = z^n. \quad (37)$$

La derivada de esta función $w' = nz^{n-1}$ se anula para $z = 0$. La transformación (37) conviene considerarla utilizando coordenadas polares o la forma trigonométrica de un número complejo. Sea

$$z = r(\cos \phi + i \sin \phi),$$

$$w = \varrho(\cos \theta + i \sin \theta).$$

Utilizando el hecho de que en la multiplicación de números complejos los módulos se multiplican y los argumentos se suman, obtenemos

$$z^n = r^n (\cos n\phi + i \sin n\phi),$$

y, por tanto,

$$\varrho = r^n,$$

$$\theta = n\phi.$$

En la última fórmula vemos que el rayo $\phi = \text{const.}$ del plano z se transforma en el rayo $\theta = n\phi = \text{const.}$ en el plano w . Por tanto, un ángulo α entre dos rayos en el plano z se transformará en un ángulo de magnitud $\beta = n\alpha$. La transformación del plano z en el plano w deja de ser «uno a uno». En efecto, un punto dado w de módulo ϱ

y argumento θ se puede obtener como imagen de cada uno de los n puntos de módulo $r = \sqrt[n]{\varrho}$ y argumentos

$$\phi = \frac{\theta}{n}, \frac{\theta}{n} + \frac{2\pi}{n}, \dots, \frac{\theta}{n} + \frac{2\pi}{n}(n-1).$$

Al elevar a la potencia n , los módulos de los correspondientes puntos serán iguales a ϱ y sus argumentos iguales a

$$\theta, \theta + 2\pi, \dots, \theta + 2\pi(n-1),$$

y como al variar el valor del argumento en un múltiplo de 2π no cambia la posición geométrica del punto, todas las imágenes sobre el plano w son idénticas.

Transformaciones conformes

Si una función analítica $w = f(z)$ transforma «uno a uno» un dominio D del plano z en un dominio Δ del plano w , entonces decimos que dicha función efectúa una transformación conforme del dominio D en el dominio Δ .

La importancia de las transformaciones conformes en la teoría de funciones y sus aplicaciones se debe al siguiente teorema casi trivial.

Si $\zeta = F(w)$ es una función analítica en el dominio Δ , entonces la función compuesta $F[f(z)]$ es una función analítica sobre el dominio D . Este teorema resulta de la ecuación

$$\frac{\Delta\zeta}{\Delta z} = \frac{\Delta\zeta}{\Delta w} \cdot \frac{\Delta w}{\Delta z}.$$

Puesto que las funciones $\zeta = F(w)$ y $w = f(z)$ son analíticas, llegamos a la conclusión de que los dos factores del segundo miembro tienen un límite, y por tanto en cada punto del dominio D el cociente $\Delta\zeta/\Delta z$ tiene un límite único $d\zeta/dz$. Esto prueba que la función $\zeta = F[f(z)]$ es analítica.

El teorema que acabamos de demostrar muestra que el estudio de funciones analíticas sobre el dominio Δ se puede reducir al estudio de funciones analíticas sobre el dominio D . Si la estructura geométrica de D es más sencilla, este hecho simplifica el estudio de las funciones.

La clase más importante de dominios en la que es necesario estudiar funciones analíticas es la de los dominios *simplemente conexos*, nombre que se da a los dominios cuya frontera consiste en una sola pieza (Fig. 15a), en oposición a aquellos cuya frontera consta de varias piezas (por ejemplo, los dominios ilustrados en las figuras 15b y 15c).

Resaltemos que a veces interesa estudiar funciones sobre un dominio exterior, no interior, a una curva. Si el contorno del mismo consta de una sola pieza, entonces el dominio se llama también simplemente conexo (Fig. 15d.)

En la base de la teoría de transformaciones conformes se encuentra el siguiente teorema de Riemann.

Para un dominio simplemente conexo arbitrario Δ , es posible construir una función analítica que efectúe una transformación conforme del círculo de radio unidad y centro en el origen en el dominio dado, de tal forma que

el centro del círculo se transforme en un punto dado w_0 del dominio Δ , y una curva con una dirección arbitraria en el centro del círculo se transforme en una curva con una dirección arbitraria en el punto w_0 . Este teorema muestra que el estudio de funciones de variable compleja sobre dominios simplemente conexos arbitrarios se puede reducir al estudio de funciones definidas, por ejemplo, sobre el círculo unidad.

Expondremos ahora un bosquejo general de cómo se pueden aplicar estos resultados a problemas de la teoría del ala de avión. Supongamos que deseamos estudiar el flujo en torno a un perfil curvado de forma arbitraria.

Si logramos construir una transformación conforme del dominio exterior al perfil en el dominio exterior al círculo unidad, entonces podemos hacer uso de la función característica del flujo alrededor del círculo para construir la función característica del flujo alrededor del perfil.

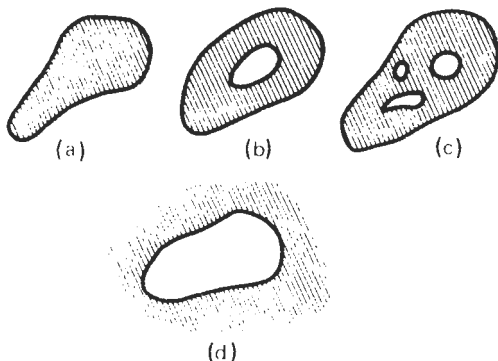


FIG. 15

Sea ζ el plano del círculo, z el plano del perfil, y $\zeta = f(z)$ una función que efectúa la transformación del dominio exterior al perfil en el dominio exterior al círculo, donde

$$\lim_{z \rightarrow \infty} \zeta = \infty.$$

Denotemos por a el punto del círculo correspondiente al borde del perfil A y construyamos el flujo circulatorio alrededor del círculo, de modo que una de las líneas de corriente abandone el círculo en a (Fig. 16). Esta función se denotará por $W(\zeta)$:

$$W(\zeta) = \Phi + i\Psi.$$

Las líneas de corriente de este flujo están definidas por la ecuación

$$\Psi = \text{const.}$$

Consideremos ahora la función

$$w(z) = W[f(z)],$$

y pongamos

$$w = \phi + i\psi.$$

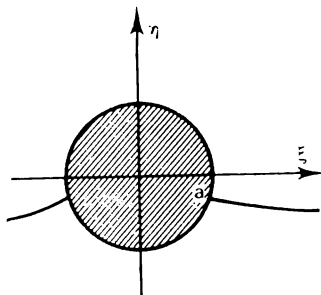


FIG. 16

Probemos que $w(z)$ es la función característica del flujo exterior al perfil con una línea de corriente que abandona el perfil en el punto A . Antes de nada veamos que el flujo definido por la función $w(z)$ es realmente un flujo exterior al perfil. Para demostrar esto, debemos probar que el contorno del perfil es una línea de corriente, es decir, que sobre el contorno del perfil

$$\psi(x, y) = \text{const.}$$

Pero esto se deduce del hecho de que

$$\psi(x, y) = \Psi(\xi, \eta),$$

y los puntos (x, y) que están sobre el perfil corresponden a los puntos (ξ, η) que están sobre el círculo, donde $\Psi(\xi, \eta) = \text{const.}$

También es fácil probar que A es un punto de estancamiento del flujo, y se puede demostrar que mediante una elección adecuada de la velocidad para el flujo exterior al círculo podemos obtener un flujo exterior al perfil con cualquier velocidad deseada.

El importante papel que juegan las transformaciones conformes en la teoría de funciones y sus aplicaciones dio origen a numerosos problemas de hallar la transformación conforme de un dominio en otro con una forma geométrica dada. En una serie de casos sencillos, pero útiles, este problema se puede resolver mediante funciones elementales. Pero en el caso general las funciones elementales no son suficientes. Como ya vimos, el teorema general de la teoría de transformaciones conformes fue enunciado por Riemann, aunque él no dió una demostración rigurosa. De hecho, una demostración completa requirió el esfuerzo conjunto de muchos grandes matemáticos durante un período de varias décadas.

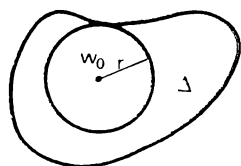


FIG. 17

En estrecha conexión con las diferentes tentativas para demostrar el teorema de Riemann surgieron métodos de aproximación para la construcción general de transformaciones conformes de dominios. La construcción efectiva de la transformación conforme de un dominio en otro es a veces un problema muy difícil. Para el estudio de muchas de las propiedades generales de las funciones no es necesario, con frecuencia, conocer la transformación efectiva de un dominio sobre otro, sino que basta con aprovechar algunas de sus propiedades geométricas, lo cual ha llevado a un amplio estudio de las propiedades geométricas de las transformaciones conformes. Para ilustrar la naturaleza de los teoremas de esta clase formularemos uno de ellos.

Transformemos el círculo de radio uno del plano z y centro el origen en cierto dominio dado (Fig. 17). Si consideramos una transformación completamente arbitraria del círculo en el dominio Δ no podemos asegurar nada acerca de su comportamiento en el punto $z = 0$. Pero para transformaciones conformes tenemos el teorema siguiente.

La dilatación en el origen no es superior a cuatro veces el radio del círculo de centro en w_0 , inscrito en el dominio

$$|f'(0)| \leq 4r.$$

A lo largo de un gran número de estudios, los matemáticos soviéticos consideraron distintas cuestiones de la teoría de transformaciones conformes. En estos trabajos se han obtenido fórmulas exactas para muchas clases interesantes de transformaciones conformes, se han desarrollado métodos para el cálculo aproximado de las mismas y se han establecido muchos teoremas geométricos generales sobre este tipo de transformaciones.

Transformaciones cuasi-conformes

Las transformaciones conformes están estrechamente conectadas con el estudio de las funciones analíticas, es decir, con el estudio de un par de funciones que verifican las condiciones de Cauchy-Riemann

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y},$$

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Pero en muchos problemas de física matemática intervienen sistemas de ecuaciones diferenciales más generales, que pueden estar también relacionados con transformaciones de un plano en otro, y estas transformaciones tendrán propiedades geométricas específicas en un entorno de los puntos del plano Oxy . Para ilustrarlo, consideremos el siguiente ejemplo de ecuaciones diferenciales

$$\frac{\partial u}{\partial x} = p(x, y) \frac{\partial v}{\partial y},$$

$$\frac{\partial v}{\partial x} = -p(x, y) \frac{\partial u}{\partial y} \quad (38)$$

Cuando $p(x, y) = 1$, este sistema de ecuaciones es el de Cauchy-Riemann. En el caso general de una función arbitraria $p(x, y)$, podemos considerar también cada solución del sistema (38) como una transformación del plano Oxy en el plano Ouv . Examinemos las propiedades geométricas de esta transformación en el entorno de un punto (x_0, y_0) . Tomando un entorno pequeño de (x_0, y_0) , nos quedamos sólo con los primeros términos del desarrollo de las fun-

ciones u y v en potencias de $x - x_0$ e $y - y_0$, y consideramos así la siguiente transformación afín

$$\begin{aligned} u - u_0 &= \left(\frac{\partial u}{\partial x} \right)_0 (x - x_0) + \left(\frac{\partial u}{\partial y} \right)_0 (y - y_0), \\ v - v_0 &= \left(\frac{\partial v}{\partial x} \right)_0 (x - x_0) + \left(\frac{\partial v}{\partial y} \right)_0 (y - y_0). \end{aligned} \quad (39)$$

Si las funciones u y v verifican el sistema de ecuaciones (38), entonces para esta transformación afín tenemos la siguiente propiedad.

Las elipses con centro en el punto (x_0, y_0) , ejes principales paralelos a los ejes de coordenadas, y con razón de semejejes

$$\frac{b}{a} = p(x_0, y_0)$$

se transforman en el plano Ouv en círculos con centro en el punto (u_0, v_0) .

Probemos esta proposición. La ecuación del círculo con centro (u_0, v_0) en el plano Ouv es

$$(u - u_0)^2 + (v - v_0)^2 = \varrho^2.$$

Sustituyendo $u - u_0$ y $v - v_0$ por sus expresiones en términos de x e y , obtenemos la ecuación para la curva correspondiente en el plano Oxy :

$$\begin{aligned} & \left[\left(\frac{\partial u}{\partial x} \right)_0^2 + \left(\frac{\partial v}{\partial x} \right)_0^2 \right] (x - x_0)^2 + \\ & + 2 \left[\left(\frac{\partial u}{\partial x} \right)_0 \left(\frac{\partial u}{\partial y} \right)_0 + \left(\frac{\partial v}{\partial x} \right)_0 \left(\frac{\partial v}{\partial y} \right)_0 \right] (x - x_0) (y - y_0) + \\ & + \left[\left(\frac{\partial u}{\partial y} \right)_0^2 + \left(\frac{\partial v}{\partial y} \right)_0^2 \right] (y - y_0)^2 = \varrho^2. \end{aligned}$$

Utilizando las ecuaciones (38) para expresar las derivadas de v en función de las de u , obtenemos

$$\left[\left(\frac{\partial u}{\partial x} \right)_0^2 + p^2 \left(\frac{\partial u}{\partial y} \right)_0^2 \right] (x - x_0)^2 + \frac{1}{p^2} \left[\left(\frac{\partial u}{\partial x} \right)_0^2 + p^2 \left(\frac{\partial u}{\partial y} \right)_0^2 \right] (y - y_0)^2 = \varrho^2$$

Si ponemos

$$a = \frac{\varrho}{\sqrt{\left(\frac{\partial u}{\partial x} \right)_0^2 + p^2 \left(\frac{\partial u}{\partial y} \right)_0^2}},$$

$$b = \frac{p\varrho}{\sqrt{\left(\frac{\partial u}{\partial x} \right)_0^2 + p^2 \left(\frac{\partial u}{\partial y} \right)_0^2}},$$

esta ecuación toma la forma

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1.$$

Por tanto, la curva que se transforma en un círculo es, en efecto, una elipse con las propiedades indicadas.

Si, en vez de considerar la transformación afín dada por los primeros términos del desarrollo, trabajamos, por el contrario, con la transformación exacta, entonces la propiedad anterior de la transformación será tanto más exacta cuanto menores sean las elipses, por lo que podemos decir que la propiedad se verifica para elipses infinitamente pequeñas.

De esta manera, de las ecuaciones (38) se deduce que en cada punto los semiejes de la elipse infinitesimal que se transforma en un círculo vienen completamente determinados por la transformación, tanto en lo que toca a su dirección como a la razón de sus longitudes. Se puede probar que esta propiedad geométrica caracteriza completamente el sistema de ecuaciones diferenciales (38); es decir, si las funciones u y v efectúan una transformación con la propiedad geométrica dada, entonces satisfacen este sistema de ecuaciones. De esta manera el problema de estudiar las soluciones de las ecuaciones (38) es equivalente a estudiar transformaciones con las propiedades dadas.

Observemos, en particular, que para las ecuaciones de Cauchy-Riemann esta propiedad se formula de la manera siguiente.

Un círculo infinitesimal con centro en el punto (x_0, y_0) se transforma en un círculo infinitesimal con centro en el punto (u_0, v_0) .

Una clase muy amplia de ecuaciones de física matemática se puede reducir al estudio de transformaciones con las siguientes propiedades geométricas.

Para cada punto (x, y) del plano del argumento nos dan la dirección de los semiejes de dos elipses y también la razón de las longitudes de estos semiejes. Deseamos construir una transformación del plano Oxy en el Ouv tal que una elipse infinitesimal de la primera familia se transforme en una elipse infinitesimal de la segunda con centro en el punto (u, v) .

Las transformaciones relacionadas con tales sistemas generales de ecuaciones fueron introducidas por el matemático soviético M. A. Laurentiev y han recibido el nombre de *cuasi-conformes*. La idea de estudiar transformaciones definidas mediante sistemas de ecuaciones diferenciales permitió extender los métodos de la teoría de funciones analíticas a una amplia clase de problemas. Laurentiev y sus discípulos desarrollaron el estudio de las transformaciones cuasi-conformes y hallaron un gran número de aplicaciones a distintos problemas de física matemática, mecánica y geometría. Es interesante observar que el estudio de transformaciones cuasi-conformes ha demostrado ser muy fructífero en la teoría de funciones analíticas. Por supuesto, no podemos profundizar aquí en las distintas aplicaciones del método geométrico a la teoría de funciones de variable compleja.

§4. La integral de línea; fórmula de Cauchy y sus corolarios

Integrales de funciones de variable compleja

En el estudio de las propiedades de funciones analíticas, el concepto de variable compleja desempeña un papel muy importante. En correspondencia con la integral definida de una función de variable real, trataremos aquí de la integral de una función de una variable compleja a lo largo de una curva. Consideremos en el plano una curva C con origen en el punto z_0 y extremo en z , y una función $f(z)$ definida en un dominio que contenga a la curva C . Dividamos la curva C en pequeños segmentos (Fig. 18) por medio de los puntos

$$z_0, z_1, \dots, z_n = z$$

y consideremos la suma

$$S = \sum_{k=1}^n f(z_k) (z_k - z_{k-1}).$$

Si la función $f(z)$ es continua y la curva C tiene longitud finita, podemos probar, igual que para funciones reales, que cuando el número de puntos de división aumenta y la distancia entre dos puntos consecutivos decrece hacia cero, la suma S tiende a un límite completamente determinado. Este límite se denomina *integral a lo largo de la curva C* y se denota por

$$\int_C f(z) dz.$$

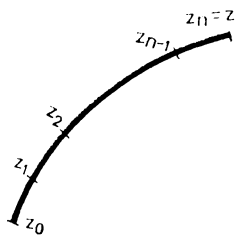


FIG. 18

Observemos que en esta definición de integral hemos distinguido entre el origen y el extremo de la curva C ; en otras palabras, hemos elegido un sentido específico de movimiento sobre la curva C .

Es fácil probar algunas propiedades sencillas de la integral.

1. La integral de la suma de dos funciones es igual a la suma de las integrales de cada una de ellas:

$$\int_C [f(z) + g(z)] dz = \int_C f(z) dz + \int_C g(z) dz.$$

2. Los factores constantes se pueden sacar fuera del signo integral:

$$\int_C A f(z) dz = A \int_C f(z) dz.$$

3. Si la curva C es suma de las curvas C_1 y C_2 , entonces

$$\int_C f(z) dz = \int_{C_1} f(z) dz + \int_{C_2} f(z) dz.$$

4. Si $\bar{C}$ es la curva C con orientación opuesta, entonces

$$\int_{\bar{C}} f(z) dz = - \int_C f(z) dz.$$

Todas estas propiedades son evidentes para las sumas aproximadas y se extienden a la integral en el paso al límite.

5. Si la longitud de la curva C es igual a L y si, en todo punto de C se verifica la desigualdad

$$|f(z)| \leq M,$$

entonces

$$\left| \int_C f(z) dz \right| \leq ML.$$

Demostremos esta propiedad. Basta probar la desigualdad para la suma S , puesto que entonces se conservará también para la integral por paso al límite. Para la suma, tenemos:

$$|S| = \left| \sum f(z_k)(z_k - z_{k-1}) \right| \leq \sum |f(z_k)| |z_k - z_{k-1}| \leq M \sum |z_k - z_{k-1}|.$$

Pero la suma del segundo factor es igual a la suma de las longitudes de los segmentos de la poligonal inscrita en la curva C con vértices en los puntos z_k . La longitud de la poligonal, como ya sabemos, no es mayor que la longitud de la curva, por lo que

$$|S| \leq ML.$$

Consideremos la integral de la función más sencilla $f(z) = 1$. Evidentemente, en este caso

$$S = (z_1 - z_0) + (z_2 - z_1) + \cdots + (z_n - z_{n-1}) = z_n - z_0 = z - z_0.$$

Esto prueba que

$$\int_C 1 \cdot dz = z - z_0.$$

Este resultado nos muestra que para la función $f(z) = 1$ el valor de la integral para todas las curvas que unen los puntos z_0 y z es el

mismo. En otras palabras, el valor de la integral depende solamente de los extremos del camino de integración. Pero es fácil probar que esta propiedad no es válida para funciones arbitrarias de una variable compleja. Por ejemplo, si $f(z) = x$, entonces un cálculo sencillo muestra que

$$\int_{C_1} x dz = \frac{x^2}{2} + i y x, \quad \int_{C_2} x dz = \frac{x^2}{2}, \quad z = x + iy,$$

donde C_1 y C_2 son los caminos de integración que se indican en la figura 19.

Dejamos al lector que compruebe estas ecuaciones.

Un hecho notable en la teoría de funciones analíticas nos lo da el siguiente teorema de Cauchy.

Si $f(z)$ es diferenciable en cada punto de un dominio simplemente conexo D , entonces las integrales a lo largo de todos los caminos que unen dos puntos arbitrarios z_0 y z del dominio son iguales.

No daremos aquí una demostración del teorema de Cauchy; para ello remitimos al lector interesado a cualquier libro de teoría de funciones de variable compleja. Mencione-

mos algunas consecuencias importantes de este teorema.

Ante todo, el teorema de Cauchy nos permite introducir la integral indefinida de una función analítica. Fijemos el punto z_0 y consideremos la integral a lo largo de curvas que unan z_0 y z :

$$F(z) = \int_{z_0}^z f(\zeta) d\zeta.$$

Aquí podemos tomar la integral sobre cualquier curva que una z_0 con z , ya que al cambiar la curva no varía el valor de la integral, la cual, por tanto, depende solamente de z . La función $F(z)$ es una *integral indefinida* de $f(z)$.

Una integral indefinida de $f(z)$ tiene una derivada igual a $f(z)$.

En muchas aplicaciones conviene tener una formulación ligeramente distinta del teorema de Cauchy, como sigue.

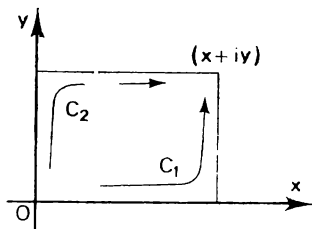


FIG. 19

Si $f(z)$ es diferenciable en cada punto de un dominio simplemente conexo, entonces la integral a lo largo de cualquier contorno cerrado contenido en dicho dominio es igual a cero:

$$\int_{\Gamma} f(z) dz = 0.$$

Esto es evidente, ya que en un contorno cerrado los dos extremos coinciden, por lo que z_0 y z se pueden unir por un camino nulo.

Al hablar de contorno cerrado nos referimos a una curva recorrida en el sentido contrario a las agujas de un reloj. Si el contorno se recorre en el sentido del reloj la denotaremos por $\bar{\Gamma}$.

La integral de Cauchy

Apoyándonos en el último teorema podemos probar la siguiente fórmula fundamental de Cauchy que expresa el valor de una función diferenciable en puntos interiores a un contorno cerrado mediante los valores de la función sobre la curva:

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{\zeta - z}.$$

A continuación damos una demostración de esta fórmula. Sea z fijo y ζ una variable independiente. La función

$$\phi(\zeta) = \frac{f(\zeta)}{\zeta - z}$$

será continua y diferenciable en cada punto ζ interior al dominio D , a excepción del punto $\zeta = z$, donde se anula el denominador, circunstancia que impide la aplicación del teorema de Cauchy a la función $\phi(\zeta)$ sobre la curva C .

Consideremos un círculo K_ϱ , con centro en el punto z y radio ϱ , y demostremos que

$$\int_C \phi(\zeta) d\zeta = \int_{K_\varrho} \phi(\zeta) d\zeta. \quad (40)$$

Para ello construimos el contorno cerrado auxiliar Γ_ϱ , formado por el contorno C , el camino γ_ϱ , que conecta C con el círculo, y el círculo

$\bar{K}_e$, tomado con la orientación opuesta (Fig. 20). El contorno Γ_e está indicado por flechas. Como el punto $\zeta = z$ está excluido, la función $\phi(\zeta)$ es diferenciable en todo punto interior a Γ_e , y por tanto,

$$\int_{\Gamma_e} \phi(\zeta) d\zeta = 0. \quad (41)$$

Pero el contorno Γ_e está dividido en cuatro partes: C , γ_e , $\bar{K}_e$ y $\bar{\gamma}_e$, por lo que, en virtud de la propiedad 3 de la última subsección, tenemos

$$\int_{\Gamma_e} \phi(\zeta) d\zeta = \int_C \phi(\zeta) d\zeta + \int_{\gamma_e} \phi(\zeta) d\zeta + \int_{\bar{K}_e} \phi(\zeta) d\zeta + \int_{\bar{\gamma}_e} \phi(\zeta) d\zeta = 0.$$

Sustituyendo las integrales a lo largo de $\bar{K}_e$ y $\bar{\gamma}_e$ por integrales a lo largo de K_e y γ_e , y utilizando la propiedad 4, obtendremos

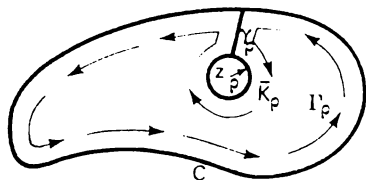


FIG. 20

$$\begin{aligned} \int_{\Gamma_e} \phi(\zeta) d\zeta &= \\ &= \int_C \phi(\zeta) d\zeta - \int_{K_e} \phi(\zeta) d\zeta = 0, \end{aligned}$$

lo cual prueba la fórmula (40)

Para calcular el segundo miembro de (40), ponemos

$$\begin{aligned} \int_{K_e} \phi(\zeta) d\zeta &= \int_{K_e} \frac{f(\zeta)}{\zeta - z} d\zeta = \int_{K_e} \frac{f(\zeta) - f(z)}{\zeta - z} d\zeta + \int_{K_e} \frac{f(z) d\zeta}{\zeta - z} = \\ &= \int_{K_e} \frac{f(\zeta) - f(z)}{\zeta - z} d\zeta + f(z) \int_{K_e} \frac{d\zeta}{\zeta - z}. \end{aligned} \quad (42)$$

Calculemos primero el segundo término. Sobre el círculo K_e ,

$$\zeta = z + \varrho(\cos \theta + i \sin \theta).$$

Utilizando el hecho de que z y ϱ son constantes, obtenemos

$$d\zeta = \varrho(-\sin \theta + i \cos \theta) d\theta = i\varrho(\cos \theta + i \sin \theta) d\theta,$$

y por tanto,

$$\zeta - z = \varrho(\cos \theta + i \operatorname{sen} \theta),$$

de donde

$$\int_{\kappa_\varrho} \frac{d\zeta}{\zeta - z} = \int_{\kappa_\varrho} i d\theta = 2\pi i,$$

ya que, para una vuelta completa de la circunferencia, la variación total de θ es igual a 2π . De (40) y (42) tenemos

$$\int_C \frac{f(\zeta) d\zeta}{\zeta - z} = 2\pi i f(z) + \int_{\kappa_\varrho} \frac{f(\zeta) - f(z)}{\zeta - z} d\zeta.$$

Tomemos límites en esta ecuación para $\varrho \rightarrow 0$. El primer miembro y el primer término del segundo miembro permanecerán invariables. Si demostramos que el límite del segundo término es igual a cero, entonces nuestra ecuación nos da la fórmula de Cauchy. Para probar que el segundo término tiende a cero cuando $\varrho \rightarrow 0$, observemos que

$$\lim_{\zeta \rightarrow z} \frac{f(\zeta) - f(z)}{\zeta - z} = f'(\zeta),$$

es decir, la expresión bajo el signo integral tiene un límite finito, y por tanto está acotada

$$\left| \frac{f(\zeta) - f(z)}{\zeta - z} \right| < M.$$

Aplicando la propiedad 5 de la integral, tenemos

$$\left| \int_{\kappa_\varrho} \frac{f(\zeta) - f(z)}{\zeta - z} d\zeta \right| \leq M 2\pi \varrho \rightarrow 0.$$

Esto completa la demostración de la fórmula de Cauchy. Esta es una de las herramientas básicas de la investigación en la teoría de funciones de variable compleja.

Desarrollo de funciones diferenciables en serie de potencias

Aplicamos el teorema de Cauchy para establecer dos propiedades básicas de las funciones diferenciables de una variable compleja.

Toda función de una variable compleja que tenga derivada primera en un dominio D tiene derivadas de todos los órdenes.

En efecto, en el interior de un contorno cerrado nuestra función se puede expresar mediante la fórmula integral de Cauchy

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{\zeta - z} d\zeta.$$

La función de z bajo el signo de integración es diferenciable; por tanto, derivando bajo el signo integral, obtenemos

$$f'(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta - z)^2} d\zeta.$$

La función que aparece bajo el signo integral es también diferenciable; por consiguiente, podemos derivar otra vez, obteniendo

$$f''(z) = \frac{1 \cdot 2}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{(\zeta - z)^3}.$$

Continuando la derivación, obtenemos la fórmula general

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{(\zeta - z)^{n+1}}$$

De esta forma podemos calcular la derivada de cualquier orden. Para que esta demostración sea completamente rigurosa habría que probar también que la derivación bajo el signo integral es válida. No daremos esta parte de la demostración.

La segunda propiedad es la siguiente:

Si $f(z)$ es diferenciable en todos los puntos de un círculo K con centro en el punto a , entonces $f(z)$ puede desarrollarse en una serie de Taylor

$$f(z) = f(a) + \frac{f'(a)}{1!} (z - a) + \cdots + \frac{f^{(n)}(a)}{n!} (z - a)^{n+1} + \cdots,$$

que converge en el interior de K .

En la sección 1 hemos definido las funciones analíticas de una variable compleja como funciones desarrollables en serie de poten-

cias. Este último teorema dice que toda función diferenciable de una variable compleja es analítica. Esta es una propiedad especial de las funciones de variable compleja que no tiene análoga en el campo real. Una función de variable real que tenga derivada primera puede no tener derivada segunda en algún punto.

Probemos el teorema formulado en el párrafo anterior.

Supongamos que $f(z)$ es derivable en el interior y sobre la frontera del círculo K con centro en el punto a . Entonces en el interior de K la función $f(z)$ se puede expresar mediante la integral de Cauchy

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{\zeta - z}. \quad (43)$$

Escribimos

$$\zeta - z = (\zeta - a) - (z - a);$$

entonces

$$\frac{1}{\zeta - z} = \frac{1}{(\zeta - a) - (z - a)} = \frac{1}{\zeta - a} \frac{1}{1 - \frac{z - a}{\zeta - a}} \quad (44)$$

Utilizando el hecho de que el punto z es interior al círculo y que ζ está sobre la circunferencia, obtenemos

$$\left| \frac{z - a}{\zeta - a} \right| < 1,$$

de modo que, aplicando la fórmula básica para una serie geométrica, tenemos:

$$\frac{1}{1 - \frac{z - a}{\zeta - a}} = 1 + \left(\frac{z - a}{\zeta - a} \right) + \cdots + \left(\frac{z - a}{\zeta - a} \right)^n + \cdots, \quad (45)$$

y la serie de la derecha converge. Utilizando (44) y (45), podemos representar la fórmula (43) en la forma

$$f(z) = \frac{1}{2\pi i} \int_C \left[\frac{f(\zeta)}{\zeta - a} + (z - a) \frac{f(\zeta)}{(\zeta - a)^2} + \cdots + \right. \\ \left. + (z - a)^n \frac{f(\zeta)}{(\zeta - a)^{n+1}} + \cdots \right] d\zeta.$$

Integremos ahora término a término la serie entre corchetes. (La validez de esta operación se puede establecer rigurosamente.) Sacando el factor $(z - a)^n$, que no depende de ζ , fuera del signo integral en cada término, obtenemos

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{\zeta - a} + \frac{z - a}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{(\zeta - a)^2} + \cdots + \\ + \frac{(z - a)^n}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{(\zeta - a)^{n+1}} + \cdots.$$

Utilizando ahora las fórmulas integrales para la sucesión de derivadas, podemos escribir

$$\frac{1}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{(\zeta - a)^{n+1}} = \frac{f^{(n)}(a)}{n!},$$

de donde obtenemos

$$f(z) = f(a) + \frac{f'(a)}{1!} (z - a) + \cdots + \frac{f^{(n)}(a)}{n!} (z - a)^n + \cdots.$$

Hemos probado que las funciones diferenciables de variable compleja se pueden desarrollar en serie de potencias. Recíprocamente, las funciones representadas mediante series de potencias son diferenciables. Sus derivadas se pueden calcular derivando término a término la serie. (La validez de esta operación se puede establecer rigurosamente.)

Funciones enteras

Una serie de potencias da una representación analítica de una función solamente en un círculo. Este círculo tiene un radio igual a la distancia al punto más próximo en el que la función deja de ser analítica, es decir, al punto singular más próximo de la función.

Entre las funciones analíticas es natural distinguir la clase de funciones que son analíticas para todos los valores finitos de la variable. Tales funciones están representadas por series de potencias, convergentes para todos los valores de variable z , y se denominan *funciones enteras* de z . Si consideramos desarrollos en torno al origen,

entonces una función entera vendrá expresada mediante una serie de la forma

$$G(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n + \cdots$$

Si en esta serie todos los coeficientes, a partir de uno, son iguales a cero, entonces la función es simplemente un polinomio, o una función racional entera

$$P(z) = c_0 + c_1 z + \cdots + c_n z^n.$$

Si en el desarrollo hay infinitos términos que son distintos de cero, entonces la función entera se denomina *trascendente*.

Ejemplos de tales funciones son:

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots,$$

$$\operatorname{sen} z = \frac{z}{1!} - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots,$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots.$$

En el estudio de las propiedades de los polinomios juega un papel importante la distribución de las raíces de la ecuación

$$P(z) = 0,$$

o, hablando con más generalidad, podemos plantear el problema de la distribución de los puntos para los cuales el polinomio tiene un valor dado A

$$P(z) = A.$$

El teorema fundamental del álgebra dice que todo polinomio toma un valor dado A al menos en un punto. Esta propiedad no se puede extender a una función entera arbitraria. Por ejemplo, la función $w = e^z$ no toma el valor cero en ningún punto del plano z . Sin embargo, tenemos el siguiente teorema de Picard: toda función entera toma cada valor arbitrariamente prefijado una infinidad de veces, con la posible excepción de un valor.

La distribución de los puntos del plano en los que una función

entera toma un valor dado A constituye una de las cuestiones centrales de la teoría de funciones enteras.

El número de raíces de un polinomio es igual a su grado. El grado de un polinomio está estrechamente relacionado con la rapidez de crecimiento de $|P(z)|$ cuando $|z| \rightarrow \infty$. En efecto, podemos escribir

$$|P(z)| = |z|^n \cdot \left| a_n + \frac{a_{n-1}}{z} + \cdots + \frac{a_0}{z^n} \right|,$$

y como para $|z| \rightarrow \infty$ el segundo factor tiende a $|a_n|$, un polinomio de grado n crece —para grandes valores de $|z|$ — como $|a_n| \cdot |z|^n$. Por tanto, está claro que para valores mayores de n , el crecimiento de $|P_n(z)|$ para $|z| \rightarrow \infty$ será más rápido y el polinomio tendrá también más raíces. Este principio también es válido para funciones enteras. Sin embargo, para una función entera, en general, existen infinitas raíces, y por tanto la cuestión del número de raíces no tiene sentido. No obstante, podemos considerar el número de raíces $n(r, a)$ de la ecuación

$$f(z) = a$$

en un círculo de radio r , e investigar cómo varía este número cuando aumenta r . El grado de crecimiento de $n(r, a)$ resulta estar relacionado con el ritmo de crecimiento del máximo $M(r)$ del módulo de la función entera sobre el círculo de radio r . Como ya enunciamos anteriormente, para una función entera puede existir un valor excepcional de a para el que la ecuación puede no tener ni siquiera una raíz. Para cualquier otro valor de a , el ritmo de crecimiento del número $n(r, a)$ es comparable al ritmo de crecimiento de la cantidad $\ln M(r)$. No podemos dar aquí formulaciones más exactas de estas leyes.

Las propiedades de la distribución de las raíces de funciones enteras están relacionadas con problemas de la teoría de números, y han permitido a los matemáticos establecer muchas propiedades importantes de las funciones enteras de Riemann⁶, a partir de las cuales es posible demostrar muchos teoremas sobre números primos.

Funciones fraccionarias o meromorfas

La clase de funciones enteras se puede considerar como una extensión de la clase de los polinomios algebraicos. A partir de los

⁶ Véase el capítulo 10 sobre la teoría de números.

polinomios podemos obtener una clase más amplia de funciones racionales

$$R(z) = \frac{P(z)}{Q(z)},$$

constituidas por cocientes de dos polinomios.

De manera análoga es natural formar una nueva clase de funciones a partir de las funciones enteras. El cociente de dos funciones enteras $G_1(z)$ y $G_2(z)$

$$f(z) = \frac{G_1(z)}{G_2(z)},$$

se denomina función *fraccionaria* o *meromorfa*. La clase que se origina de esta manera desempeña un gran papel en análisis matemático. Entre las funciones elementales contenidas en la clase de funciones meromorfas están, por ejemplo:

$$\operatorname{tg} z = \frac{\operatorname{sen} z}{\cos z}, \quad \operatorname{cotg} z = \frac{\cos z}{\operatorname{sen} z}.$$

Una función meromorfa no será analítica en todo el plano complejo. En aquellos puntos en que el denominador $G_2(z)$ se anule, la función $f(z)$ se hace infinita. Las raíces de $G_2(z)$ forman un conjunto de puntos aislados en el plano. En un entorno de estos puntos, $f(z)$ no se puede, naturalmente, desarrollar en serie de Taylor; sin embargo, en un entorno de uno de ellos, a , una función meromorfa se puede representar mediante una serie de potencias que contiene también un cierto número de potencias negativas de $(z - a)$:

$$f(z) = \frac{C_{-m}}{(z-a)^m} + \cdots + \frac{C_{-1}}{z-a} + C_0 + C_1(z-a) + \cdots + C_n(z-a)^n + \cdots \quad (46)$$

Cuando z se aproxima al punto a , el valor de $f(z)$ tiende a infinito. Un punto singular aislado en el que una función analítica se hace infinita se denomina *polo*. La pérdida de analiticidad de la función en el punto a proviene de los términos con potencias negativas de $z - a$ en el desarrollo (46). La expresión

$$\frac{C_{-m}}{(z-a)^m} + \cdots + \frac{C_{-1}}{z-a}$$

caracteriza el comportamiento de una función meromorfa cerca de un punto singular y se denomina la *parte principal* del desarrollo (46). En un entorno de un polo, el comportamiento de una función meromorfa está determinado por su parte principal. En muchos casos, si conocemos la parte principal del desarrollo de una función meromorfa en un entorno de cada uno de sus polos, podemos construir la función. Así, por ejemplo, si $f(z)$ es racional y se anula en el infinito, entonces es igual a la suma de las partes principales de sus desarrollos en torno a todos sus polos, cuyo número, para una función racional, es finito:

$$f(z) = \sum_{(k)} \left[\frac{C_{-m_k}^{(k)}}{(z - a_k)^{m_k}} + \cdots + \frac{C_{-1}^{(k)}}{z - a_k} \right].$$

En el caso general una función racional, se puede representar como la suma de todas sus partes principales y un polinomio:

$$f(z) = \sum_{(k)} \left[\frac{C_{-m_k}^{(k)}}{(z - a_k)^{m_k}} + \cdots + \frac{C_{-1}^{(k)}}{z - a_k} \right] + C_0 + C_1 z + \cdots + C_m z^m \quad (47)$$

La fórmula (47) da, para una función racional, una expresión en la que pone de manifiesto el papel representado por sus puntos singulares. La expresión (47) es de gran utilidad en diversas aplicaciones de las funciones racionales, pero también tiene un gran interés teórico desde el momento en que muestra de qué manera los puntos singulares de la función determinan su estructura total. Ocurre que, igual que para una función racional, toda función meromorfa se puede construir a partir de las partes principales de sus polos. Introducimos sin demostrar la expresión apropiada, por ejemplo, para la función $\cotg z$. Los polos de esta función son las raíces de la ecuación

$$\operatorname{sen} z = 0$$

y están situados en los puntos: $\cdots, -k\pi, \cdots, -\pi, 0, \pi, \cdots, k\pi, \cdots$. Se puede demostrar que la parte principal del desarrollo de la función $\cotg z$ en serie de potencias en el polo $z = k\pi$ es

$$\frac{1}{z - k\pi},$$

y la función $\cotg z$ es igual a la suma de las partes principales con respecto a todos los polos

$$\cotg z = \frac{1}{z} + \sum_{k=1}^{\infty} \left(\frac{1}{z - k\pi} + \frac{1}{z + k\pi} \right). \quad (48)$$

El desarrollo de una función meromorfa en una serie de las partes principales es notable por mostrar claramente la posición de todos los puntos singulares y porque nos permite calcular la función sobre todo el dominio de definición.

La teoría de funciones meromorfas ha llegado a ser fundamental para el estudio de muchas clases de funciones que son de gran importancia en análisis. En particular, debemos subrayar su significación para las ecuaciones de la física matemática. La creación de la teoría de ecuaciones integrales, que responde a muchas cuestiones de la teoría de ecuaciones de física matemática, se basó en grado sumo en los teoremas fundamentales relativos a funciones meromorfas.

Desde entonces, el desarrollo de la parte del análisis funcional que guarda una relación más estrecha con la física matemática, a saber, la teoría de operadores, ha dependido muy a menudo de hechos relativos a la teoría de funciones analíticas.

Sobre la representación analítica de funciones

Anteriormente vimos que en un entorno de cada punto donde una función es diferenciable ésta se puede definir mediante una serie de potencias. Para una función entera, la serie de potencias converge en todo el plano y da una expresión analítica de la función dondequiera que esté definida. Cuando la función no es entera, la serie de Taylor, como sabemos, converge solamente en un círculo cuya circunferencia pasa por el punto singular más próximo de la función. Por consiguiente, la serie no nos permite calcular la función en todos los puntos, y así puede suceder que una función analítica no se pueda representar mediante una serie de potencias en la totalidad de su dominio de definición. Para una función meromorfa, una expresión analítica que especifica la función sobre todo el dominio de definición es el desarrollo en partes principales.

Si una función no es entera pero está definida en algún círculo, o bien si tenemos una función definida en algún dominio pero queremos estudiarla solamente en un círculo, entonces la serie de Taylor puede servir para representarla. Pero, cuando estudiamos la función

en dominios que no son círculos, surge la cuestión de hallar para la función una expresión analítica adecuada que la represente en todo el dominio. Los términos de una serie de potencias que expresa una función analítica en un círculo son los polinomios más sencillos $a_n z^n$. Es natural preguntar si, dentro de un dominio arbitrario, podemos desarrollar una función analítica en una serie de polinomios más general. Entonces cada término de la serie se podría calcular mediante operaciones aritméticas, y obtendríamos un método para representar funciones que estaría basado una vez más en las operaciones más simples de la aritmética. La respuesta general a esta cuestión la da el siguiente teorema.

Una función analítica, dada en un dominio arbitrario cuyo contorno consista en una sola curva, se puede desarrollar en una serie de polinomios

$$f(z) = P_0(z) + P_1(z) + \cdots + P_n(z) + \cdots$$

El teorema que acabamos de formular da sólo una respuesta general a la cuestión de desarrollar una función en una serie de polinomios en un dominio arbitrario, pero no nos permite, sin embargo, construir la serie para una función dada, como se hizo anteriormente en el caso de la serie de Taylor. Este teorema plantea, más que resuelve, la cuestión de desarrollar funciones en una serie de polinomios. El problema de construir la serie de polinomios a partir de la función o de alguna de sus propiedades, el de construir series que converjan más rápidamente o series estrechamente relacionadas con el comportamiento de la función, el problema de la estructura de una función definida mediante una serie de polinomios, todos ellos representan un extenso desarrollo de la teoría de aproximación de funciones mediante series de polinomios. En la creación de la misma han representado un gran papel los matemáticos soviéticos, que han obtenido diversos resultados fundamentales.

§5. *Propiedades de unicidad y prolongación analítica*

Propiedades de unicidad de las funciones analíticas

Una de las propiedades más notables de las funciones analíticas es su unicidad, como expresa el siguiente teorema.

Dadas dos funciones analíticas en un dominio D, si coinciden sobre una curva C interior al dominio, entonces coinciden en todo él.

La demostración de este teorema es muy sencilla. Sean $f_1(z)$ y $f_2(z)$ dos funciones analíticas en el dominio D y tales que coinciden sobre la curva C . La diferencia

$$\phi(z) = f_1(z) - f_2(z)$$

será una función analítica sobre el dominio D y se anulará sobre C . Probemos ahora que $\phi(z) = 0$ en todo punto de D . En efecto, si en D existe un punto z_0 (Fig. 21) en el cual $\phi(z_0) \neq 0$, prolongamos la curva C hasta el punto z_0 y proseguimos a lo largo de la curva prolongada hasta z_0 , mientras la función permanezca igual a cero sobre Γ . Sea ζ el último punto de Γ que es accesible de esta forma. Si $\phi(z_0) \neq 0$, entonces $\zeta \neq z_0$, y sobre un segmento de Γ a partir de ζ la función $\phi(z)$, por la definición del punto ζ , no será igual a cero. Probemos que esto es imposible. En efecto, sobre la parte Γ_ζ de la curva Γ anterior al punto ζ , tenemos $\phi(z) = 0$. Podemos calcular todas las derivadas de $\phi(z)$ sobre Γ_ζ utilizando solamente los valores de $\phi(z)$ sobre Γ_ζ , de modo que sobre Γ_ζ todas las derivadas de $\phi(z)$ son iguales a cero. En particular, en el punto ζ

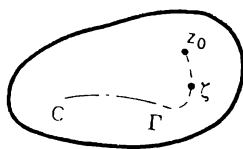


FIG. 21

$$\phi(\zeta) = \phi'(\zeta) = \cdots = \phi^{(n)}(\zeta) = \cdots = 0.$$

Desarrollemos $\phi(z)$ en serie de Taylor en el punto ζ . Todos los coeficientes del desarrollo son nulos, de modo que obtenemos

$$\phi(z) = 0$$

en algún círculo con centro en el punto ζ , contenido en el dominio D . En particular, se deduce que la ecuación $\phi(z) = 0$ se debe verificar en algún segmento de la curva Γ posterior a ζ . La hipótesis $\phi(z_0) \neq 0$ nos lleva a una contradicción.

Este teorema prueba que, si conocemos los valores de una función analítica en un segmento de una curva o sobre una parte de un dominio, entonces los valores de la función están unívocamente determinados en todos los puntos del dominio dado. Por consiguiente, los valores de una función analítica en distintas partes del plano están estrechamente relacionados unos con otros.

Para comprender el significado de esta propiedad de unicidad de

una función analítica sólo es necesario recordar que la definición general de función de una variable compleja permite cualquier ley de correspondencia entre valores de la variable y de la función. Con tal definición ni siquiera podemos plantearnos el determinar los valores de una función en cualquier punto mediante sus valores en otra parte del plano. Vemos que el mero requisito de la diferenciabilidad de una función de una variable compleja es tan fuerte que determina la relación entre valores de la función en distintos puntos.

Hagamos también hincapié en que en la teoría de funciones de una variable real la diferenciabilidad de una función no lleva por sí misma a ninguna consecuencia similar. En efecto, podemos construir ejemplos de funciones que son diferenciables infinitas veces y que coinciden en alguna parte del eje Ox pero difieren en otra. Por ejemplo, una función igual a cero para todo valor negativo de x se puede definir de tal manera que para x positivos sea diferente de 0 y tenga derivadas continuas de cualquier orden. Para ello basta, por ejemplo, poner, para $x > 0$

$$f(x) = e^{-1/x}.$$

Prolongación analítica y funciones analíticas completas

El dominio de definición de una función de variable compleja viene restringido a menudo por la manera misma de definir la función. Consideremos un ejemplo muy elemental. Sea la función dada mediante la serie

$$f(z) = 1 + z + z^2 + \cdots + z^n + \cdots. \quad (49)$$

Esta serie, como es sabido, converge en el círculo unidad y diverge fuera de él. Por tanto, la función analítica dada por la fórmula (49) está definida solamente en este círculo. Por otra parte, sabemos que la suma de la serie (49) en el círculo $|z| < 1$ viene expresada por la fórmula

$$f(z) = \frac{1}{1-z}. \quad (50)$$

La fórmula (50) tiene sentido para todo valor $z \neq 1$. Del teorema de unicidad se deduce que la expresión (50) representa la única función analítica que coincide con la suma de la serie (49) en el círculo $|z| < 1$. Así, esta función, dada en principio en el círculo unidad, ha sido extendida a todo el plano.

Si tenemos una función $f(z)$ definida en el interior de cierto dominio D , y si existe otra $F(z)$ definida en un dominio Δ que contenga a D y que coincida con $f(z)$ en D , entonces, por el teorema de unicidad, el valor de $F(z)$ en Δ está determinado de manera única.

La función $F(z)$ se llama la *prolongación analítica* de $f(z)$. Una función analítica se llama *completa* si no se puede prolongar analíticamente más allá del dominio sobre el que ya está definida. Por ejemplo, una función entera, definida para todo el plano, es completa. Una función meromorfa es también una función completa; está definida en todas partes excepto en sus polos. Sin embargo, existen funciones analíticas cuyo dominio completo de definición está acotado. No daremos aquí estos ejemplos más complicados.

El concepto de función analítica completa conduce a la necesidad de considerar funciones multiformes de una variable compleja. Ilustremos esto con el ejemplo de la función

$$\operatorname{Ln} z = \ln r + i\phi,$$

donde $r = |z|$ y $\phi = \arg z$. Si en un punto $z_0 = r_0(\cos \phi_0 + i \operatorname{sen} \phi_0)$ del plano z consideramos cierto valor inicial de la función

$$(\operatorname{Ln} z)_0 = \ln r_0 + i\phi_0,$$

entonces nuestra función analítica se puede prolongar continuamente a lo largo de una curva C . Como ya hemos mencionado anteriormente, es fácil ver que, si el punto z describe un camino cerrado C_0 , que parte del punto z_0 , da una vuelta alrededor del origen (Fig. 22) y vuelve a z_0 , entonces en este punto encontramos el valor original de $\ln r_0$, pero el ángulo ϕ ha aumentado en 2π . Esto prueba que, si prolongamos la función $\operatorname{Ln} z$ de una manera continua a lo largo del camino C , aumentamos su valor en $2\pi i$ en una vuelta del contorno C . Si el punto z se mueve a lo largo de este contorno n veces, entonces en lugar del valor original

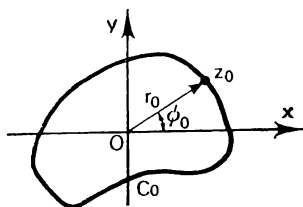


FIG. 22

$$(\operatorname{Ln} z)_0 = \ln r_0 + i\phi_0$$

obtenemos el nuevo valor

$$(\operatorname{Ln} z)_n = \ln r_0 + (2\pi n + \phi_0)i.$$

Si el punto z describe el contorno m veces en el sentido opuesto, obtenemos

$$(\operatorname{Ln} z)_{-m} = \ln r_0 + (-2\pi m + \phi_0)i.$$

Estas observaciones muestran que sobre el plano complejo estamos ineludiblemente obligados a considerar la conexión entre los distintos valores de $\operatorname{Ln} z$. Esta función tiene infinitos valores. Con respecto a su carácter multiforme, el punto $z = 0$ desempeña un papel especial, ya que en torno suyo pasamos de un valor de la función a otro. Es fácil establecer que si z describe un contorno cerrado que no rodea al origen, el valor de $\operatorname{Ln} z$ no cambia. El punto $z = 0$ se llama un punto de ramificación de la función $\operatorname{Ln} z$.

En general, si para una función $f(z)$, y a lo largo de un circuito en torno al punto a , pasamos de uno de sus valores a otro, entonces a se llama un *punto de ramificación* de la función $f(z)$.

Consideremos un segundo ejemplo. Sea

$$w = \sqrt[n]{z}.$$

Como previamente hicimos notar, esta función es también multiforme y toma n valores

$$\begin{aligned} & \sqrt[n]{r} \left(\cos \frac{\phi}{n} + i \operatorname{sen} \frac{\phi}{n} \right), \quad \sqrt[n]{r} \left(\cos \frac{\phi + 2\pi}{n} + i \operatorname{sen} \frac{\phi + 2\pi}{n} \right), \\ & \dots, \sqrt[n]{r} \left(\cos \frac{\phi + 2\pi(n-1)}{n} + i \operatorname{sen} \frac{\phi + 2\pi(n-1)}{n} \right). \end{aligned}$$

Todos los valores de nuestra función se pueden deducir del valor

$$w_0 = \sqrt[n]{r_0} \left(\cos \frac{\phi_0}{n} + i \operatorname{sen} \frac{\phi_0}{n} \right)$$

sin más que describir una curva cerrada alrededor del origen, ya que en cada vuelta en torno al origen el ángulo ϕ se incrementa en 2π .

Describiendo la curva cerrada $(n-1)$ veces, obtenemos, a partir del primer valor de $\sqrt[n]{z}$, los $(n-1)$ valores restantes. Al recorrer el contorno por n -ésima vez volvemos al valor

$$\sqrt[n]{z_0} = \sqrt[n]{r_0} \left(\cos \frac{\phi_0 + 2\pi n}{n} + i \operatorname{sen} \frac{\phi_0 + 2\pi n}{n} \right) =$$

$$= \sqrt[n]{r_0} \left(\cos \frac{\phi_0}{n} + i \operatorname{sen} \frac{\phi_0}{n} \right),$$

es decir, volvemos al valor original de la raíz.

Superficies de Riemann para funciones multiformes

Existe una manera geométrica, fácilmente visualizable, de representar el carácter de una función multiforme.

Consideremos otra vez la función $\operatorname{Ln} z$, y hagamos un corte en el plano z a lo largo de la parte positiva del eje Ox . Si impedimos que el punto z cruce el corte, entonces no podemos pasar continuamente de un valor de $\operatorname{Ln} z$ a otro. Si prolongamos $\operatorname{Ln} z$ desde el punto z_0 , solamente podemos llegar al mismo valor de $\operatorname{Ln} z$.

La función uniforme hallada de esta manera en el plano z cortado es una *rama uniforme* de la función $\operatorname{Ln} z$. Todos los valores de $\operatorname{Ln} z$ están distribuidos en un conjunto infinito de ramas uniformes

$$\ln r + i\phi, \quad 2\pi n < \phi \leq 2\pi(n+1).$$

Es fácil probar que la n -ésima rama toma el mismo valor sobre el borde inferior del corte que la rama $(n+1)$ -ésima en el borde superior.

Para distinguir las diferentes ramas de $\operatorname{Ln} z$, imaginemos infinitos ejemplares del plano z , cada uno de ellos cortado a lo largo de la parte positiva del eje Ox , y apliquemos sobre la n -ésima hoja los valores de la variable z correspondientes a la n -ésima rama. Los puntos situados en distintos ejemplares del plano pero que tienen las mismas coordenadas corresponderán a uno y al mismo número $x + iy$; pero el hecho de que este número pertenezca a la n -ésima hoja indica que estamos considerando la n -ésima rama del logaritmo.

Para representar geoméricamente el hecho de que la n -ésima rama del logaritmo, sobre la parte inferior del corte del n -ésimo plano, coincide con la rama $(n+1)$ -ésima sobre la parte superior del corte en el plano $(n+1)$ -ésimo, pegamos los planos n y $(n+1)$ -ésimos, uniendo la parte inferior del corte del primero con la superior del corte del segundo. Esta construcción nos lleva a una superficie de múltiples hojas, que tiene la forma de una escalera de caracol (Fig. 23). El punto $z = 0$ representa el papel de la columna central de la escalera.

Si un punto pasa de una hoja a otra, entonces el número complejo vuelve a su valor original, pero la función $\operatorname{Ln} z$ pasa de una rama a otra.

La superficie así construida se llama la *superficie de Riemann* de la función $\text{Ln } z$. Riemann fue el primero en introducir la idea de construir superficies que representaran el carácter de funciones analíticas multiformes, demostrando lo fructífero de esta idea.

Discutamos también la construcción de la superficie de Riemann de la función $w = \sqrt{z}$. Esta función es biforme y tiene un punto de ramificación en el origen.

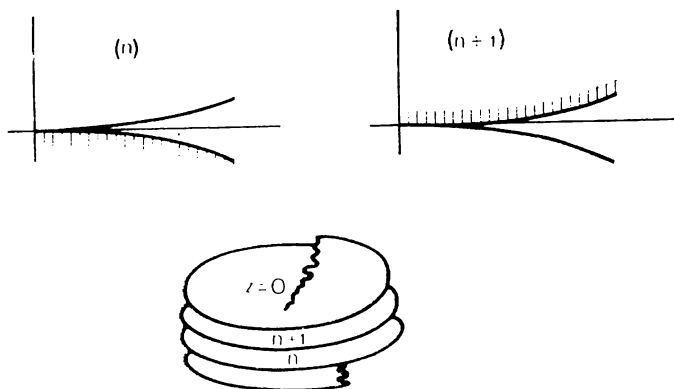


FIG. 23

Imaginemos dos ejemplares del plano z colocados uno encima de otro y ambos cortados a lo largo de la parte positiva del eje Ox . Si z parte de z_0 y describe un contorno cerrado C conteniendo al origen, entonces $\sqrt{z}$ pasa de una rama a la otra, y por tanto el punto sobre la superficie de Riemann pasa de una hoja a la otra. Para conseguir esto, pegamos el borde inferior del corte de la primera hoja al borde superior del corte de la segunda. Si z describe el contorno cerrado C por segunda vez, entonces $\sqrt{z}$ debe volver a su valor original, por lo que el punto sobre la superficie de Riemann debe volver a su posición original sobre la primera hoja. Para lograr esto, debemos unir ahora el borde inferior de la segunda hoja al borde superior de la primera. Como resultado de ello obtenemos una superficie de dos hojas, que se corta a sí misma a lo largo de la parte positiva del eje Ox . Una idea de esta superficie puede obtenerse de la figura 24, que muestra las proximidades del punto $z = 0$.

De la misma forma podemos construir una superficie de muchas hojas para representar el carácter de una función multiforme cual-

quiera. Las distintas hojas de tal superficie están conectadas unas con otras en torno a puntos de ramificación de la función. Sucede que las propiedades de las funciones analíticas están estrechamente relacionadas con las propiedades geométricas de las superficies de Riemann. Estas superficies no sólo son un medio auxiliar para ilustrar el carácter de una función multiforme, sino que también desempeñan un papel fundamental en el estudio de las propiedades de funciones analíticas y en el desarrollo de métodos para su investigación. Las

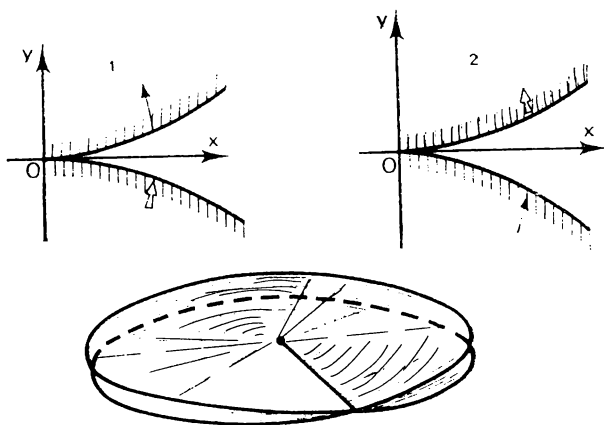


FIG. 24

superficies de Riemann proporcionaron una especie de puente entre el análisis y la geometría en el campo de las variables complejas, que permite no sólo relacionar con la geometría las propiedades analíticas más profundas de las funciones, sino también desarrollar un campo totalmente nuevo de la geometría: la topología, que estudia aquellas propiedades geométricas de las figuras que permanecen invariantes bajo deformaciones continuas.

Uno de los ejemplos más claros del significado de las propiedades geométricas de las superficies de Riemann es la teoría de funciones algebraicas, es decir, funciones obtenidas como soluciones de una ecuación

$$f(z, w) = 0,$$

cuyo primer miembro es un polinomio en z y w . La superficie de Riemann de una función tal se puede siempre deformar continuamente

en una esfera o, si no, en una esfera con asas (Fig. 25). La propiedad característica de estas superficies es el número de asas. Este número se llama género de la superficie y de la función algebraica correspondiente. Sucede que el género de una función algebraica determina sus propiedades más importantes.

§6. Conclusión

La teoría de funciones analíticas surgió en relación con el problema de resolución de ecuaciones algebraicas. Pero, a medida que fue desarrollándose, se relacionó constantemente con ramas cada vez

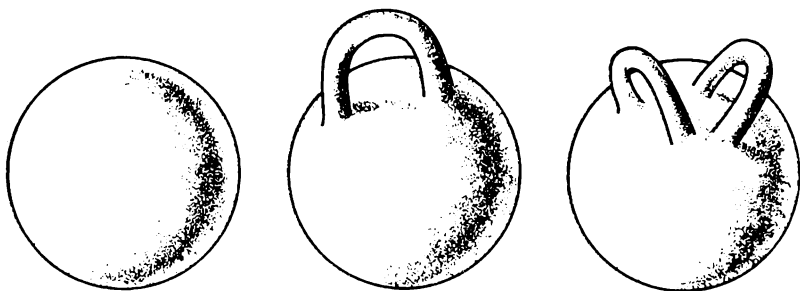


FIG. 25

más nuevas de las matemáticas. Arrojó luz sobre las clases fundamentales de funciones del análisis, la mecánica y la física matemática. Muchos de los hechos centrales del análisis pudieron aclararse en último término pasando al campo complejo. Las funciones de variable compleja recibieron una interpretación física inmediata en los importantes campos vectoriales de la hidrodinámica y la electrodinámica, y proporcionaron una notable herramienta para la resolución de los problemas que surgen en estas ramas de la ciencia. Se descubrieron relaciones entre la teoría de funciones y los problemas de la teoría de propagación del calor, elasticidad, etc.

Las cuestiones generales de la teoría de ecuaciones diferenciales y métodos especiales para su resolución se han basado siempre, en gran parte, sobre la teoría de funciones de variable compleja. Las funciones analíticas se introdujeron de una manera natural en la

teoría de ecuaciones integrales y en la teoría general de operadores lineales. Además, se descubrieron estrechas relaciones entre la teoría de funciones analíticas y la geometría. Todas estas conexiones, cada vez más amplias, de la teoría de funciones con nuevas áreas de las matemáticas y otras ciencias, muestran la vitalidad de la teoría y el continuo enriquecimiento de su gama de problemas.

En esta panorámica general no hemos podido presentar un cuadro completo de las múltiples ramificaciones de la teoría de funciones. Hemos intentado solamente dar una idea de la naturaleza sumamente variada de sus problemas, indicando los hechos elementales básicos de las distintas direcciones fundamentales en las cuales se ha movido la teoría. Algunos de sus aspectos más importantes, su conexión con la teoría de ecuaciones diferenciales y funciones especiales, con funciones elípticas y automorfas, con la teoría de series trigonométricas y con otras muchas ramas de las matemáticas, han sido completamente ignorados en nuestra discusión. En otros casos hemos tenido que restringirnos a brevísimas indicaciones. Pero esperamos que este examen dará al lector una idea general del carácter y significado de la teoría de funciones de variable compleja.

Parte 4

Capítulo 10

NUMEROS PRIMOS

§1. *El estudio de la teoría de números*

Números enteros

Como el lector sabe por la introducción al capítulo 1, el género humano tuvo que habérselas, desde los tiempos más remotos, con los números enteros; pero fue necesario el transcurso de muchos siglos para que surgiera el concepto de la sucesión infinita de números naturales

$$1, 2, 3, 4, 5, \dots, \quad (1)$$

Hoy día, en las cuestiones más variadas de la actividad cotidiana nos enfrentamos constantemente con problemas en los que intervienen números enteros. Los números enteros reflejan numerosas relaciones cuantitativas de la naturaleza; en todas las cuestiones que tienen conexión con objetos discretos, constituyen el aparato matemático necesario.

Por otra parte, los números enteros desempeñan un papel importante en el estudio de continuos. Así, por ejemplo, en análisis matemático se considera el desarrollo de una función analítica en una serie de potencias enteras de x

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n + \dots$$

Todos los cálculos son esencialmente llevados a cabo con números *enteros*, como se ve inmediatamente tras un examen superficial de las máquinas calculadoras automáticas, de las calculadoras de mesa o de las tablas matemáticas, tales como las de logaritmos. Después de efectuadas estas operaciones con números enteros, se colocan las comas decimales en posiciones bien definidas, correspondientes a la formación de fracciones decimales; éstas, como todas las fracciones racionales, representan cocientes de dos números *enteros*. Cuando trabajamos con un número *real* cualquiera en trabajos prácticos (por ejemplo, π) lo reemplazamos en realidad por una fracción racional (por ejemplo, suponemos que $\pi = 22/7$, o que $\pi = 3,14$).

Mientras que el establecimiento de las reglas para operar con los números concierne a la aritmética, las propiedades más profundas de la sucesión (1) de números naturales —extendida para incluir el cero y los números negativos— se estudian en la *teoría de números*, que es la ciencia del sistema de los enteros y, en un sentido más amplio, también de los sistemas de números contruidos de una manera determinada a partir de los enteros (véase, en particular, la sección 5 de este capítulo). Bien entendido que la teoría de números considera los enteros, no separados unos de otros, sino como interdependientes; la teoría de números estudia aquellas propiedades de los enteros que están definidas por ciertas relaciones entre ellos.

Una de las cuestiones básicas en la teoría de números concierne a la *divisibilidad* de un número por otro; si el resultado de dividir el entero a por el entero b (distinto de cero) es un entero, es decir, si

$$a = b \cdot c$$

(a , b y c enteros) entonces decimos que a es *divisible* por b o que b *divide* a a . Si el resultado de dividir el entero a por el entero b es una fracción, entonces decimos que a no es divisible por b . Cuestiones relativas a la divisibilidad de números surgen constantemente en la práctica y juegan también un importante papel en algunos problemas del análisis matemático. Por ejemplo, si el desarrollo de una función en potencias enteras de x

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots \quad (2)$$

es tal que todos los coeficientes impares (con subíndices no divisibles por 2) son iguales a cero, es decir, si

$$f(x) = a_0 + a_2x^2 + \cdots + a_{2k}x^{2k} + \cdots,$$

entonces la función verifica la condición

$$f(-x) = f(x);$$

una función tal se dice que es par, y su gráfica es simétrica con respecto al eje de ordenadas. Pero, si en el desarrollo (2) todos los coeficientes pares (con subíndices divisibles por 2) son nulos, es decir, si

$$f(x) = a_1x + a_3x^3 + \cdots + a_{2k+1}x^{2k+1} + \cdots,$$

entonces

$$f(-x) = -f(x);$$

en este caso la función se llama impar, y su gráfica es simétrica con respecto al origen.

Así, por ejemplo, tenemos que:

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \quad (\text{función impar});$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \quad (\text{función par}).$$

La cuestión geométrica de la posibilidad de construir un n -ágono regular con regla y compás depende de la naturaleza aritmética del número n ¹.

Un *número primo* es un entero (mayor que uno) que tiene solamente dos divisores enteros positivos: 1 y él mismo. El 1 no se considera como primo, ya que no tiene dos divisores positivos diferentes.

Así, pues, los números primos son

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \cdots \quad (3)$$

Los números primos desempeñan un papel fundamental en la teoría de números debido al siguiente teorema básico: todo entero $n > 1$

¹ Véase el capítulo 4.

se puede representar como producto de números primos (con posible repetición de factores), es decir, en la forma

$$n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}, \quad (4)$$

donde $p_1 < p_2 < \cdots < p_k$ son primos y $a_1, a_2, \cdots, a_k$ son enteros no menores que uno; además, la representación de n en la forma (4) es única.

Las propiedades relacionadas con la representación de números como una suma de términos se llaman *aditivas*; las relativas a su representación en forma de un producto se llaman *multiplicativas*. La conexión entre las propiedades aditivas y multiplicativas de los números es extraordinariamente complicada y ha dado origen a una serie de problemas en la teoría de números.

La existencia de estos difíciles problemas en la teoría de números, junto con el hecho de que el número entero no sólo es el más sencillo y claro de todos los conceptos matemáticos, sino que está estrechamente relacionado con la realidad objetiva, ha conducido a la creación —con vistas a su aplicación a la teoría de números— de nuevas ideas muy profundas y de potentes métodos, muchos de los cuales han llegado a tener también importancia en otras ramas de las matemáticas. Por ejemplo, la idea de la sucesión infinita de números naturales, que refleja la infinitud del mundo material en espacio y tiempo, ha ejercido una gran influencia sobre todos los desarrollos de las matemáticas. También tiene un gran significado el hecho de que los términos de la sucesión de números naturales estén ordenados. El estudio de las operaciones con enteros ha conducido al concepto de operación algebraica, que desempeña un papel capital en diversas ramas de las matemáticas.

De inmensa importancia para las matemáticas ha sido el concepto, particularmente útil en la aritmética, de algoritmo, procedimiento de resolución de problemas basado en la repetición reiterada de un proceso estrictamente definido; en particular, el papel del algoritmo es fundamental para el uso de máquinas matemáticas. La naturaleza esencial del método algorítmico para resolver problemas está claramente ilustrada por el algoritmo de Euclides para hallar el máximo común divisor de dos números naturales a y b .

Supongamos que $a > b$. Dividimos a por b y hallamos el cociente q_1 ; si b no divide a a , habrá un resto r_2 :

$$a = bq_1 + r_2, \quad 0 < r_2 < b. \quad (5_1)$$

Además, si $r_2 \neq 0$, dividimos b por r_2

$$b = r_2 q_2 + r_3, \quad 0 < r_3 < r_2. \quad (5_2)$$

Luego dividimos r_2 por r_3 y continuamos el proceso hasta obtener un resto nulo, que debe existir necesariamente para una sucesión decreciente de enteros no negativos $r_2, r_3, \dots$. Sea

$$r_{n-2} = r_{n-1} q_{n-1} + r_n, \quad (5_{n-1})$$

$$r_{n-1} = r_n q_n; \quad (5_n)$$

entonces r_n es el máximo común divisor de a y b . Porque, si dos enteros l y m tienen un divisor común d , entonces para cualesquiera enteros h y k el número $hl + km$ será también divisible por d . Denotemos el máximo común divisor de a y b por δ . Por la ecuación (5_1) vemos que δ es divisor de r_2 ; de (5_2) se deduce que δ es divisor de $r_3, \dots$; de (5_{n-1}) , que δ es divisor de r_n . Pero r_n es un divisor común a a y b , ya que en (5_n) vemos que r_n divide a r_{n-1} ; en (5_{n-1}) , que r_n divide a r_{n-2} , etc. Por tanto, δ es idéntico a r_n , y el problema de hallar el máximo común divisor de a y b está resuelto. Tenemos así un procedimiento bien definido, del mismo tipo para todo a y b , que nos lleva automáticamente al resultado deseado y es, por consiguiente, un ejemplo característico de algoritmo.

La teoría de números ha ejercido su influencia sobre el desarrollo de muchas disciplinas matemáticas: análisis matemático, geometría, álgebra clásica y moderna, teoría de sumabilidad de series, teoría de probabilidades, etc.

Métodos de la teoría de números

En cuanto a sus métodos, la teoría de números se divide en cuatro partes: elemental, analítica, algebraica y geométrica.

La teoría elemental de números estudia las propiedades de los enteros sin recurrir a otras disciplinas matemáticas. Así, partiendo de la identidad de Euler

$$\begin{aligned} (x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = & (x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4)^2 + \\ & + (x_1 y_2 - x_2 y_1 + x_3 y_4 - x_4 y_3)^2 + (x_1 y_3 - x_3 y_1 + x_4 y_2 - x_2 y_4)^2 + \\ & + (x_1 y_4 - x_4 y_1 + x_2 y_3 - x_3 y_2)^2, \end{aligned} \quad (6)$$

podemos probar muy fácilmente que todo entero $N > 0$ se puede expresar como suma de los cuadrados de cuatro enteros; es decir, todo entero es representable en la forma

$$N = x^2 + y^2 + z^2 + u^2,$$

donde x, y, z y u son enteros ².

La teoría analítica hace uso del análisis matemático para resolver problemas de la teoría de números. Sus fundamentos fueron obra de Euler, y fue desarrollada por P. L. Chebichev, Dirichlet, Riemann, Ramanujan, Hardy, Littlewood y otros matemáticos; sus métodos más potentes se deben a Vinogradov. Esta parte de la teoría de números está estrechamente relacionada con la de funciones de una variable compleja (teoría que es muy rica en aplicaciones prácticas), y también con la de series, la teoría de probabilidades y otras ramas de las matemáticas.

El concepto fundamental de la teoría algebraica de números es el de número algebraico, es decir, raíz de la ecuación

$$a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n = 0,$$

donde $a_0, a_1, a_2, \cdots, a_n$ son enteros ³.

Las mayores contribuciones a esta rama de la teoría de números se deben a Lagrange, Gauss, Kummer, E. I. Zolotarev, Dedekind, A. O. Gel'fond y otros.

Los objetos básicos de estudio de la teoría geométrica de números son los «reticulados»; esto es, sistemas constituidos totalmente por puntos «enteros», cuyas coordenadas, en un sistema coordenado rectilíneo dado, rectangular u oblicuo, son todas ellas enteras. Los reticulados tienen gran importancia en geometría y cristalografía, y están íntimamente relacionados con importantes cuestiones de la teoría de números; en particular, con la teoría aritmética de formas cuadráticas, es decir, la teoría de formas cuadráticas con coeficientes enteros y variables enteras. Trabajos fundamentales en la teoría geométrica de números se deben a H. Minkowski y G. F. Voronoi.

Debemos señalar que los métodos de la teoría analítica de números tienen importantes aplicaciones en otras dos ramas, la algebraica

² Tenemos aquí un ejemplo de ecuación indeterminada que hay que estudiar desde el punto de vista de su resolubilidad en el campo de los enteros.

³ Si $a_0 = 1$, el número algebraico se denomina entero algebraico. Un número que no es algebraico se llama trascendente.

y la geométrica. Particularmente notable es el problema de contar el número de puntos enteros en un dominio dado, que es de gran importancia en ciertas ramas de la física. G. F. Voronoi indicó distintos medios de aproximación para este problema, e I. M. Vinogradov desarrolló métodos para su resolución.

La razón fundamental de la potencia de los métodos analíticos en la teoría de números es que enriquecen nuestro estudio de las interrelaciones entre enteros discretos al traer en nuestra ayuda nuevas relaciones entre magnitudes continuas.

Debemos recalcar que en este capítulo consideramos solamente ciertas cuestiones seleccionadas de la teoría de números.

§2. *La investigación de problemas relativos a números primos*

El número de primos es infinito

Al considerar la sucesión (3) de números primos

$$2, 3, 5, 7, 11, 13, 17, 19, \dots$$

es natural preguntarnos: ¿Esta sucesión es infinita? El hecho de que cualquier entero se pueda representar en la forma (4) todavía no resuelve el problema, ya que los exponentes $a_1, \dots, a_k$ pueden tomar un número infinito de valores. Euclides dio una respuesta afirmativa a esta pregunta, demostrando que el número de primos no puede ser igual a ningún entero finito k .

Sean $p_1, p_2, \dots, p_k$ primos; entonces el número

$$m = p_1 p_2 \cdots p_k + 1,$$

como es un entero mayor que uno, o es primo o tiene un factor primo. Pero m no es divisible por ninguno de los primos $p_1, p_2, \dots, p_k$, ya que, si lo fuera, la diferencia $m - p_1 p_2 \cdots p_k$ sería también divisible por este número, lo cual es imposible, ya que la diferencia es igual a uno. Por tanto, o m es primo o es divisible por algún primo p_{k+1} , distinto de $p_1, \dots, p_k$. Por consiguiente el conjunto de primos no puede ser finito.

La criba de Eratóstenes

El matemático griego Eratóstenes, en el siglo III a. C., describió el siguiente método de «criba» para hallar todos los primos que no

excedan de un número natural dado N . Escribimos todos los enteros desde 1 hasta N

$$1, 2, 3, 4, \dots, N,$$

y tachamos, empezando por la izquierda, primero el número 1, después todos los múltiplos de 2 excepto el 2, después todos los múltiplos de 3 excepto el 3, después todos los múltiplos de 5 salvo el 5 (los múltiplos de cuatro ya han sido tachados), y así sucesivamente; los números que queden serán entonces primos. Conviene observar que el proceso de tachar se debe continuar sólo hasta que hayamos encontrado todos los primos menores que $\sqrt{N}$, ya que todo número compuesto (es decir, no primo) que no sea mayor que N , tendrá necesariamente un divisor primo que no exceda de $\sqrt{N}$.

El examen de la sucesión de números primos dentro de la sucesión de todos los enteros positivos nos induce a pensar que la ley de distribución de números primos debe ser muy complicada; por ejemplo, encontramos primos tales como 8.004.119 y 8.004.121 (llamados primos gemelos) cuya diferencia es dos, y también primos que están lejos uno de otro, tales como 86.629 y 86.677, entre los cuales no hay ningún otro primo. Pero las tablas muestran que «por término medio» los números primos escasean cada vez más conforme avanzamos en la sucesión de enteros.

Identidad de Euler; su demostración de que el número de primos es infinito

El gran matemático del siglo XVIII, L. Euler, miembro de la Academia Rusa de Ciencias, introdujo la siguiente función, de argumento $s > 1$, que actualmente se denota por $\zeta(s)$:

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots + \frac{1}{n^s} + \dots \quad (7)$$

Como vimos en el capítulo 2, esta serie converge para $s > 1$ (y diverge para $s \leq 1$). Euler dedujo una notable identidad que desempeña un papel muy importante en la teoría de números primos:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1 - \frac{1}{p^s}}, \quad (8)$$

donde el símbolo $\prod_p$ significa que debemos multiplicar las expresiones $1/[1 - (1/p^s)]$ para todos los primos p . Para ver cómo se hace la demostración de esta identidad, observemos que $1/(1 - q) = 1 + q + q^2 + \dots$ para $|q| < 1$, de modo que

$$\frac{1}{1 - \frac{1}{p^s}} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots$$

Multiplicando estas series para los distintos primos p y recordando que todo n es representable de manera única como producto de primos, obtenemos que

$$\prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots \right) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots + \frac{1}{n^s} + \dots$$

Para una demostración rigurosa, naturalmente debemos establecer la validez de nuestro proceso de paso al límite, pero esto no presenta dificultades especiales.

De la identidad (8) podemos deducir como un corolario el hecho de que la serie $\sum_p 1/p$, formada por los inversos de todos los primos, diverge (esto proporciona una nueva demostración del hecho ya conocido de que el número de primos no puede ser finito), y también que el cociente del número de primos no superiores a x , dividido por x , converge a cero cuando x crece indefinidamente.

Las investigaciones de P. L. Chebichev sobre la distribución de los números primos en la sucesión de números naturales

Denotemos por $\pi(x)$, como es costumbre ahora, el número de primos no superiores a x ; por ejemplo, $\pi(10) = 4$, ya que 2, 3, 5 y 7 son todos los primos no superiores a 10, $\pi(\pi) = 2$, puesto que 2 y 3 son todos los primos no superiores a π . Como ya observamos anteriormente

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x} = 0.$$

Pero ¿cómo decrece la razón $\pi(x)/x$? O, dicho con otras palabras, ¿cuál es la ley de crecimiento para $\pi(x)$? ¿Podemos buscar una función conocida, razonablemente sencilla, que difiera poco de $\pi(x)$?

El famoso matemático francés Legendre, examinando tablas de números primos, estableció que tal función era

$$\frac{x}{\ln x - A}, \quad (9)$$

donde $A = 1,08 \dots$, pero no dio una demostración de esta proposición. Gauss, que también consideró la cuestión de la distribución de los números primos, conjeturó que $\pi(x)$ difería relativamente poco de $\int_2^x dt/\ln t$ (observemos que la siguiente relación es válida:

$$\lim_{x \rightarrow \infty} \frac{\int_2^x \frac{dt}{\ln t}}{\frac{x}{\ln x}} = 1, \quad (10)$$

que se establece integrando por partes y hallando estimaciones para la nueva integral).

El primer matemático desde los tiempos de Euclides que hizo progresos reales en la tan difícil cuestión de la distribución de los números primos fue P. L. Chebichev. En 1848, basando su trabajo en un estudio de la función $\zeta(s)$ de Euler para s real, Chebichev demostró que, para un n positivo arbitrariamente grande y para un α positivo arbitrariamente pequeño, existen valores arbitrariamente grandes de x para los cuales

$$\pi(x) > \int_2^x \frac{dt}{\ln t} - \frac{\alpha x}{\ln^n x}$$

y también x arbitrariamente grandes para los cuales

$$\pi(x) < \int_2^x \frac{dt}{\ln t} + \frac{\alpha x}{\ln^n x},$$

en buena concordancia con la hipótesis de Gauss. En particular tomando $n = 1$ y aplicando (10), Chebichev estableció que

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1, \quad (11)$$

siempre que el límite en (11) exista.

Chebichev refutó también la hipótesis de Legendre relativa al valor de la constante A que aparece en la expresión (9) como aquella que da la mejor aproximación a $\pi(x)$; demostró que este valor solamente podía ser $A = 1$.

El conocido matemático francés Bertrand fue llevado por sus investigaciones en la teoría de grupos a la siguiente conjetura, que comprobó empíricamente mediante tablas hasta valores muy grandes de n : Si $n > 3$, entonces entre n y $2n - 2$ hay al menos un número primo. Todos los intentos de Bertrand, y de otros matemáticos, para probar esta conjetura fueron infructuosos hasta 1850, cuando Chebichev publicó su segundo artículo sobre números primos, en el cual no sólo probaba la conjetura («postulado de Bertrand»), sino que también demostraba que para x suficientemente grande

$$A_1 < \frac{\pi(x)}{\frac{x}{\ln x}} < A_2, \quad (12)$$

donde

$$0,92 < A_1 < 1 \quad \text{y} \quad 1 < A_2 < 1,1.$$

En la sección 3 damos una presentación simplificada del método de Chebichev, que conduce, sin embargo, a resultados mucho menos precisos que los del mismo Chebichev.

Sus trabajos tuvieron una gran influencia sobre muchos matemáticos, concretamente Sylvester y Poincaré. En el transcurso de más de cuarenta años, diversos científicos se ocuparon de mejorar la desigualdad de Chebichev (12) (aumentando la constante del primer miembro de la desigualdad y disminuyendo la de la derecha), pero fueron incapaces de establecer la existencia del límite

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{\frac{x}{\ln x}}$$

(como ya indicamos antes, sabemos por el trabajo de Chebichev que si el límite existe es igual a uno).

Solamente en 1896, Hadamard, utilizando argumentos de la teoría de funciones de variable compleja, demostró que la función $\theta(x)$, introducida por Chebichev y definida por la ecuación

$$\theta(x) = \sum_{p \leq x} \ln p,$$

verifica la condición

$$\lim_{x \rightarrow \infty} \frac{\Theta(x)}{x} = 1, \quad (13)$$

de la cual es relativamente fácil obtener la relación (11) sin ninguna hipótesis adicional; ésta es la llamada ley asintótica para la distribución de primos.

El resultado (13) lo encontró Hadamard basándose en las investigaciones del famoso matemático alemán del siglo XIX Riemann, quien estudió la función $\zeta(s)$ de Euler (7) para valores complejos de la variable $s = \sigma + it$ (Chebichev había considerado esta función solamente para valores reales del argumento)⁴.

Riemann demostró que la función $\zeta(s)$, definida en el semiplano $\sigma > 1$ por la serie (7)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

tiene la propiedad de que

$$\zeta(s) = \frac{1}{s-1}$$

es una función entera trascendente (para $\sigma \leq 1$ la serie (7) deja de ser convergente, pero los valores de $\zeta(s)$ en el semiplano $\sigma \leq 1$ están definidos por prolongación analítica) (véase capítulo 9). Riemann hizo la conjetura («hipótesis de Riemann») de que las raíces de $\zeta(s)$ en la banda $0 \leq \sigma \leq 1$ tienen la parte real igual a $1/2$, es decir, están sobre la recta $\sigma = 1/2$; la cuestión de la exactitud de esta suposición está aún sin resolver.

Un paso importante en la demostración de (13) fue el establecimiento del hecho de que sobre la recta $\sigma = 1$ no hay raíces de $\zeta(s)$.

La investigación del comportamiento de $\zeta(s)$ condujo al desarrollo de una elegante teoría de funciones enteras y meromorfas, con importantes aplicaciones prácticas.

El trabajo de Vinogradov y sus discípulos en la teoría de números primos

De la ecuación (13), que en virtud de (10) se puede escribir en la forma

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{\int_2^x \frac{dt}{\ln t}} = 1, \quad (14)$$

⁴ En 1949, A. Selberg dio una demostración elemental (es decir, sin utilizar variables complejas) de la ley asintótica de distribución de primos.

surgió la cuestión del grado de exactitud con que la función $\int_2^x dt/\ln t$ representa a $\pi(x)$. Los mejores resultados en este sentido fueron hallados por N. G. Chudakov y estaban basados en el método de *sumas trigonométricas* de Vinogradov (este método se describirá en la sección 4), que también permitió a Chudakov disminuir considerablemente las cotas entre las cuales se puede hallar al menos un primo. Concretamente, había sido establecido con anterioridad que si consideramos la sucesión

$$1^{250}, 2^{250}, 3^{250}, \dots, n^{250}, (n+1)^{250}, \dots, \quad (15)$$

entonces, a partir de un $n = n_0$, debe existir, entre dos términos consecutivos cualesquiera, es decir, entre n^{250} y $(n+1)^{250}$, al menos un primo.

Observemos que, como se deduce de la fórmula del binomio

$$(n+1)^{250} - n^{250} > 250n^{249},$$

esta diferencia es muy grande. N. G. Chudakov logró sustituir la sucesión (15) por

$$1^4, 2^4, 3^4, \dots, n^4, (n+1)^4, \dots, \quad (16)$$

cuyos términos están considerablemente más próximos entre sí que los de la sucesión (15), pero que también contiene al menos un primo entre cada dos términos consecutivos, es decir, entre n^4 y $(n+1)^4$, a partir de cierto $n = n_0$. Posteriormente este resultado se ha mejorado reemplazando las cuartas potencias por cubos.

Si k y l son primos entre sí, es decir, no tienen divisor común mayor que uno, entonces una progresión aritmética de término general $kt + l$ contiene infinitos números primos. Este hecho, generalización del resultado de Euclides, fue establecido en el siglo XIX por Dirichlet. Pero ¿podemos hallar una cota que con seguridad no sea sobrepasada por el menor primo de la progresión $kt + l$? El matemático de Ленинград Ju. V. Linnik demostró la existencia de una constante absoluta C con la propiedad de que en la progresión $kt + l$ (k y l primos entre sí) existe necesariamente al menos un primo menor que k^C . De este modo, Linnik proporcionó una solución esencialmente completa del problema —plantado años atrás— del menor primo en una progresión aritmética; los investigadores posteriores solamente podrán disminuir el valor de la constante C . Linnik también llevó a cabo inves-

tigaciones muy importantes relativas a los ceros de la función $\zeta(s)$ y de funciones más generales.

Como ya dijimos anteriormente, los mejores resultados con respecto a la distribución de primos se obtuvieron con el método de Vinogradov para la estimación de sumas trigonométricas.

Una *suma trigonométrica* es una suma de la forma

$$\sum_{A < x < B} e^{2\pi i f(x)},$$

donde $f(x)$ es una función real de x , y x recorre todos los valores enteros comprendidos entre A y B , o algún subconjunto específico de estos valores, por ejemplo, los primos entre A y B . Como el módulo de $e^{2\pi i z}$ para z real es igual a uno, y el módulo de una suma no excede de la suma de los módulos de sus términos, tenemos

$$\left| \sum_{x=1}^p e^{2\pi i f(x)} \right| \leq P. \quad (17)$$

Esta estimación «trivial» se puede mejorar considerablemente en muchos casos; Vinogradov dio los pasos decisivos en este sentido. Concretamente, sea $f(x)$ un polinomio

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0.$$

Si todos los a son enteros, entonces $e^{2\pi i f(x)} = 1$ para x entero, y en este caso la estimación (17) evidentemente no se puede mejorar. Pero si no todos los $a_1, \dots, a_n$ son enteros, entonces, como demostró Vinogradov, la estimación (17) se puede afinar aproximando cualquiera de estos coeficientes mediante fracciones racionales cuyos denominadores no excedan de cierta cota (se puede probar que cualquier a comprendido entre 0 y 1 es representable en la forma $a = a/q + z$, donde a y q son enteros primos entre sí, $q \leq \tau$, $|z| \leq 1/q^\tau$ y τ es un entero prefijado mayor que 1).

La creación del método de sumas trigonométricas por Vinogradov le permitió resolver una serie de problemas muy complicados de la teoría de números. Concretamente, en 1937 resolvió un famoso problema propuesto por Goldbach, demostrando que todo número impar N suficientemente grande es representable como suma de tres primos

$$N = p_1 + p_2 + p_3. \quad (18)$$

Este problema surgió en 1742 en la correspondencia entre Euler y otro miembro de la Academia Rusa de Ciencias, C. Goldbach, y permaneció sin resolver durante casi dos siglos, pese a los esfuerzos de varios matemáticos eminentes.

Como hemos visto, la ecuación (4) muestra que los números primos juegan un papel fundamental en la representación *multiplicativa* de un número impar mediante primos. Es fácil probar, a partir de (18), que un número par suficientemente grande se puede representar como suma de no más de cuatro primos ⁵. De esta manera, el teorema de Vinogradov-Goldbach estableció una profunda conexión entre las propiedades aditivas y multiplicativas de los números.

El alcance del método de sumas trigonométricas creado por Vinogradov no está restringido a la teoría de números. Concretamente, juega un importante papel en la teoría de funciones y en la de probabilidades. Se puede obtener una idea del método de Vinogradov en la sección 4 de este capítulo.

Los lectores que estén interesados en un tratamiento más detallado pueden consultar el libro de Vinogradov *El método de sumas trigonométricas en la teoría de números*, después de una lectura preliminar de su libro *Fundamentos de la teoría de números*.

§3. Método de Chebichev

La función Θ de Chebichev y sus estimaciones

Vamos a dar una presentación simplificada del método de Chebichev para calcular el número de primos comprendidos entre límites dados. Por brevedad convenimos el uso de la siguiente notación: si B es una cantidad variable positiva que puede crecer indefinidamente, y A es otra cantidad tal que $|A|$ crece «no más rápidamente» que CB , donde C es una constante positiva (más concretamente, si existe una constante $C > 0$ tal que a partir de un instante tenemos siempre $|A|/B \leq C$), entonces escribiremos

$$A = \dot{O}(B).$$

Esto se suele leer: « A es una cantidad del orden de B .» Así, por ejemplo

$$\text{sen } x = O(1),$$

⁵ La validez de la conjetura de que todo número par suficientemente grande N se puede representar como suma de dos primos, sigue siendo hoy una cuestión abierta.

ya que en cualquier punto

$$\frac{|\operatorname{sen} x|}{1} \leq 1 ;$$

análogamente

$$5x^3 \cos 2x = O(x^3).$$

También denotaremos por $[x]$ la parte entera de x , es decir, el mayor entero no superior a x ; así, por ejemplo,

$$[\pi] = 3, \quad [5] = 5, \quad [-1,5] = -2, \quad [0,999] = 0.$$

Planteamos ahora la siguiente cuestión: sea p primo, y n un número natural, y denotemos por $n!$, como es usual, el producto $1 \cdot 2 \cdot 3 \cdot \dots \cdot n$; observemos incidentalmente que, al crecer n , el valor de $n!$ aumenta muy rápidamente. ¿Cuál es la mayor potencia a del primo p que divide a $n!$?

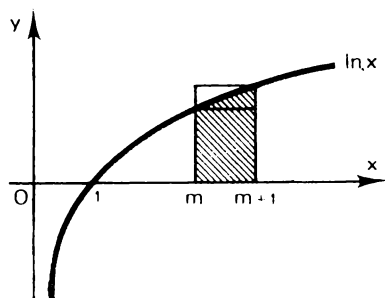


FIG. 1

Entre los números $1, 2, \dots, n$, habrá exactamente $[n/p]$ números divisibles por p ; el número de ellos que será también divisible por p^2 es $[n/p^2]$; además, de éstos habrá $[n/p^3]$ divisible por p^3 , etc. De aquí es fácil probar que

$$a = \left[\frac{n}{p} \right] + \left[\frac{n}{p^2} \right] + \left[\frac{n}{p^3} \right] + \dots$$

(la serie tiene un fin, ya que $[n/p^s] > 0$ solamente para $n \geq p^s$). Por tanto, en la última suma cada factor del producto $1 \cdot 2 \cdot 3 \cdot \dots \cdot n$ tal que la mayor potencia de p por la cual es divisible es p^m aparecerá precisamente m veces, una como múltiplo de p , otra como múltiplo de p^2 , otra como múltiplo de p^3 , $\dots$, y finalmente una como múltiplo de p^m .

De este resultado y de la representabilidad de cualquier número natural en la forma (4) se deduce que $n!$ es el producto de potencias de la forma

$$p^{\left[\frac{n}{p} \right] + \left[\frac{n}{p^2} \right] + \left[\frac{n}{p^3} \right] + \dots}$$

extendido a todos los primos $p \leq n$. Por tanto, $\ln(n!)$ será la suma de los logaritmos de estas potencias, que se puede escribir abreviadamente en la forma

$$\ln n! = \sum_{p \leq n} \left(\left[\frac{n}{p} \right] + \left[\frac{n}{p^2} \right] + \left[\frac{n}{p^3} \right] + \cdots \right) \ln p. \quad (19)$$

Simplifiquemos la ecuación (19). Como $y = \ln x$ es una función creciente, tenemos

$$\ln m = \ln m \int_m^{m+1} dx < \int_m^{m+1} \ln x dx < \ln(m+1) \int_m^{m+1} dx = \ln(m+1),$$

como se ve en la figura 1. Por consiguiente

$$\begin{aligned} \ln n! &= \ln 1 + \ln 2 + \cdots + \ln n < \\ &< \int_1^2 \ln x dx + \int_2^3 \ln x dx + \cdots + \\ &\quad + \int_{n-1}^n \ln x dx + \ln n = \int_1^n \ln x dx + \ln n; \end{aligned}$$

por otra parte

$$\begin{aligned} \ln n! &> \ln 1 + \int_1^2 \ln x dx + \cdots + \\ &\quad + \int_{n-2}^{n-1} \ln x dx + \int_{n-1}^n \ln x dx = \int_1^n \ln x dx. \end{aligned}$$

Utilizando la fórmula de integración por partes, encontramos

$$\int_1^n \ln x dx = [x \cdot \ln x]_1^n - \int_1^n x \cdot \frac{1}{x} \cdot dx = n \ln n - (n - 1).$$

Por tanto

$$n \ln n - n + 1 < \ln n! < n \ln n - n + 1 + \ln n,$$

y de aquí se deduce que

$$\ln n! = n \ln n + O(n). \quad (20)$$

Observamos que $\ln n = O(n)$; además, para $n \rightarrow \infty$, la función $\ln n$ crece más despacio que cualquier potencia positiva de n , es decir, para cualquier constante $\alpha > 0$

$$\lim_{n \rightarrow \infty} \frac{\ln n}{n^\alpha} = 0, \quad (21)$$

ya que por la regla para formas indeterminadas (véase capítulo 2)

$$\lim_{n \rightarrow \infty} \frac{\ln n}{n^\alpha} = \lim_{n \rightarrow \infty} \frac{\frac{1}{n}}{\alpha n^{\alpha-1}} = \frac{1}{\alpha} \lim_{n \rightarrow \infty} \frac{1}{n^\alpha} = 0.$$

Además, encontramos que:

$$\begin{aligned} \sum_{p \leq n} \left(\left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \right) \ln p &\leq \sum_{p \leq n} \left(\frac{n}{p^2} + \frac{n}{p^3} + \cdots \right) \ln p = \\ &= \sum_{p \leq n} \frac{n \ln p}{p^2 \left(1 - \frac{1}{p} \right)} < 2n \sum_{p \leq n} \frac{\ln p}{p^2} < 2n \sum_{m=1}^{\infty} \frac{\ln m}{m^2} = 2nC_0 = O(n), \end{aligned} \quad (22)$$

donde C_0 es la suma de la serie convergente $\sum_{m=1}^{\infty} \frac{\ln m}{m^2}$. La convergencia absoluta de esta serie se establece utilizando (21), por ejemplo, para $\alpha = 2$, mediante el criterio de comparación y el llamado criterio integral de convergencia (véase capítulo 2, sección 14). A la vista de (20) y (22), la ecuación (19) se puede poner en la forma

$$\sum_{p \leq n} \left\lfloor \frac{n}{p} \right\rfloor \ln p = n \ln n + O(n). \quad (23)$$

Consideramos ahora la función introducida por Chebichev

$$\Theta(n) = \sum_{p \leq n} \ln p \quad (24)$$

(el logaritmo del producto de todos los números primos no superiores a n).

La ecuación (23) se puede volver a escribir como:

$$\Theta\left(\frac{n}{1}\right) + \Theta\left(\frac{n}{2}\right) + \Theta\left(\frac{n}{3}\right) + \Theta\left(\frac{n}{4}\right) + \cdots = n \ln n + O(n). \quad (25)$$

En efecto, cada $\ln p$ entra en todas la sumas de la forma $\Theta(n/s)$, donde $p \leq n/s$, es decir, donde $s \leq n/p$, y el número de tales sumas $\Theta(n/s)$ es igual a $[n/p]$.

La ecuación (25) es también válida para cualquier no entero n . Para ver esto basta, evidentemente, probar que es cierto para todo x con la condición $n < x < n + 1$; y para esto es suficiente probar que sustituyendo n por x , el primer miembro de (25) no cambia, y que el primer término del segundo miembro puede aumentar en una cantidad que es $O(n)$. Pero lo primero se deduce del hecho de que tal sustitución no aumenta el valor de ninguno de los términos del primer miembro (un aumento tal sería posible solamente si n fuera incrementado en más de una unidad), y, por supuesto, el primer miembro no disminuye. Lo segundo se deduce del hecho de que por la fórmula para el incremento de una función (véase capítulo 2)

$$f(x) - f(a) = (x - a)f'(\xi), \quad a < \xi < x,$$

tenemos

$$x \ln x - n \ln n = (x - n) \cdot (\ln \xi + 1), \quad n < \xi < x,$$

y el segundo miembro de esta última ecuación es menor que $\ln(n + 1) + 1 = O(n)$, ya que $0 < x - n < 1$. De la ecuación (25) restemos dos veces la ecuación que se deduce de (25) sustituyendo n por $n/2$;

$$\begin{aligned} \Theta\left(\frac{n}{1}\right) + \Theta\left(\frac{n}{2}\right) + \Theta\left(\frac{n}{3}\right) + \Theta\left(\frac{n}{4}\right) + \cdots &= n \ln n + O(n), \\ 2\Theta\left(\frac{n}{2}\right) + 2\Theta\left(\frac{n}{4}\right) + \cdots &= 2 \cdot \frac{n}{2} \cdot \ln \frac{n}{2} + O(n) \end{aligned}$$

y obtenemos

$$\Theta\left(\frac{n}{1}\right) - \Theta\left(\frac{n}{2}\right) + \Theta\left(\frac{n}{3}\right) - \Theta\left(\frac{n}{4}\right) + \cdots = n \ln 2 + O(n) < C_1 n,$$

donde C_1 es una constante positiva. Pero $\Theta(n/1) - \Theta(n/2)$ no es mayor que todo el primer miembro, ya que las diferencias $\Theta(n/3) - \Theta(n/4)$, $\Theta(n/5) - \Theta(n/6)$, $\dots$ no pueden ser negativas. Por tanto, se deduce de esta última desigualdad que

$$\Theta\left(\frac{n}{1}\right) - \Theta\left(\frac{n}{2}\right) < C_1 n.$$

Insertando aquí los números $n/2$, $n/4$, $\dots$ en lugar de n , también obtenemos

$$\Theta\left(\frac{n}{2}\right) - \Theta\left(\frac{n}{4}\right) < C_1 \cdot \frac{n}{2},$$

$$\Theta\left(\frac{n}{4}\right) - \Theta\left(\frac{n}{8}\right) < C_1 \cdot \frac{n}{4},$$

$$\dots\dots\dots;$$

luego, utilizando el hecho de que $\Theta(n/2^k) = 0$ para k suficientemente grande (cuando $n/2^k < 2$), la adición de términos da

$$\Theta(n) < C_1 \left(n + \frac{n}{2} + \frac{n}{4} + \dots \right) = 2C_1 n. \quad (26)$$

Volviendo a la ecuación (23), hallamos

$$0 \leq \sum_{p \leq n} \frac{n}{p} \ln p - \sum_{p \leq n} \left\lfloor \frac{n}{p} \right\rfloor \ln p \leq \sum_{p \leq n} \ln p = \Theta(n) \leq 2C_1 n = O(n),$$

por lo que la ecuación (23) da

$$\begin{aligned} \sum_{p \leq n} \frac{n}{p} \ln p &= n \ln n + O(n), \\ \sum_{p \leq n} \frac{\ln p}{p} &= \ln n + \theta C, \end{aligned} \quad (27)$$

donde C es una constante mayor que cero y θ depende del número n de tal manera que $|\theta| \leq 1$.

Una estimación para el número de primos en un intervalo dado

Veamos que se puede elegir una constante positiva M de tal manera que entre n y Mn haya tantos números primos como se desee, con tal de que n sea suficientemente grande. Es decir, vamos a establecer desigualdades sencillas para el número T de primos en el intervalo $n < p \leq Mn$. Evidentemente,

$$\sum_{n < p \leq Mn} \frac{\ln p}{p} = \sum_{p \leq Mn} \frac{\ln p}{p} - \sum_{p \leq n} \frac{\ln p}{p}. \quad (28)$$

A partir de la ecuación (27), sustituyendo n por Mn , obtenemos

$$\sum_{p \leq Mn} \frac{\ln p}{p} = \ln(Mn) + \theta' C = \ln M + \ln n + \theta' C, \quad (29)$$

donde $|\theta'| \leq 1$; de este modo, a la vista de las ecuaciones (28), (29) y (27), tenemos

$$\sum_{n < p \leq Mn} \frac{\ln p}{p} = \ln M + \theta' C - \theta C = \ln M + 2\theta_0 C,$$

donde $|\theta_0| \leq 1$, es decir,

$$\ln M - 2C \leq \sum_{n < p \leq Mn} \frac{\ln p}{p} \leq \ln M + 2C. \quad (30)$$

Por otra parte, como $y = \ln x/x$ para $x > e$ es una función decreciente (ya que $y' = (1 - \ln x)/x^2 < 0$ para $\ln x > 1$, es decir, $x > e$), se deduce que para $n \geq 3$

$$T \frac{\ln Mn}{Mn} \leq \sum_{n < p \leq Mn} \frac{\ln p}{p} \leq T \frac{\ln n}{n},$$

de donde, en virtud de (30), tenemos

$$T \frac{\ln n}{n} > \ln M - 2C \quad (31)$$

y

$$T \frac{\ln(Mn)}{Mn} < \ln M + 2C, \quad (32)$$

Ahora escogemos la constante M tal que el segundo miembro de (31) sea igual a uno:

$$\ln M - 2C = 1,$$

o bien,

$$M = e^{2C+1},$$

y ponemos

$$L = M(\ln M + 2C).$$

Para el número T de primos situados entre n y Mn , obtenemos entonces, en virtud de (31) y (32), las desigualdades

$$\frac{n}{\ln n} < T < L \frac{n}{\ln n}, \quad (33)$$

que es lo que queríamos establecer. Como $n/\ln n \rightarrow \infty$ cuando n crece indefinidamente, se deduce que $T \rightarrow \infty$ también.

§4. Método de Vinogradov

Método de Vinogradov y su aplicación a la resolución del problema de Goldbach

En esta sección intentamos dar una descripción del método de Vinogradov para el caso particular del problema de Goldbach de representar un número impar como suma de tres primos.

Expresión en forma de integral para el número de representaciones de N como suma de tres primos

Sea N un número impar suficientemente grande. Notamos por $I(N)$ el número de representaciones de N como suma de tres primos; en otras palabras, el número de soluciones de la ecuación

$$N = p_1 + p_2 + p_3 \quad (34)$$

en los números primos p_1, p_2 y p_3 .

El problema de Goldbach estará resuelto si logramos establecer que $I(N) > 0$. El método de Vinogradov nos permite, no sólo establecer este hecho (para N suficientemente grande), sino también hallar una expresión aproximada para $I(N)$.

$I(N)$ se puede escribir en la siguiente forma:

$$I(N) = \sum_{p_1 \leq N} \sum_{p_2 \leq N} \sum_{p_3 \leq N} \int_0^1 e^{2\pi i(p_1 + p_2 + p_3 - N)\alpha} d\alpha, \quad (35)$$

donde las sumas están extendidas a todos los números primos no superiores a N . En efecto, para un entero $n \neq 0$

$$\int_0^1 e^{2\pi i n a} da = \frac{1}{2\pi n i} [e^{2\pi i n a}]_0^1 = \frac{1}{2\pi n i} (e^{2\pi n i} - e^0) = 0,$$

ya que

$$e^{2\pi i n} = \cos 2\pi n + i \sin 2\pi n = 1;$$

pero si $n = 0$, entonces

$$\int_0^1 e^{2\pi i n a} da = \int_0^1 da = 1.$$

Por tanto, cada vez que los primos p_1 , p_2 y p_3 suman N , la integral interior al signo sumatorio en (35) tiene el valor uno, y cuando la suma $p_1 + p_2 + p_3 \neq N$, esta integral es igual a cero, lo cual prueba la validez de la ecuación (35).

Como $e^{2\pi i a} \cdot e^{2\pi i b} = e^{2\pi i (a+b)}$ y la integral de una suma de términos es igual a la suma de las integrales de estos términos, se deduce de la ecuación (35) que

$$I(N) = \int_0^1 \left(\sum_{p \leq N} e^{2\pi i a p} \right)^3 e^{-2\pi i a N} da.$$

Introduciendo la notación

$$T_a = \sum_{p \leq N} e^{2\pi i a p} \quad (36)$$

tenemos

$$I(N) = \int_0^1 T_a^3 e^{-2\pi i a N} da. \quad (37)$$

Descomposición del intervalo de integración en intervalos básicos y complementarios

Sea h una cantidad —elegida de una manera apropiada que dependerá de N — que crece indefinidamente con N , pero que es pequeña en comparación con N e incluso con $\sqrt[3]{N/2}$, y sea $\tau = N/h$. Como la función integrada en (37) tiene un período igual a uno,

el intervalo de integración en (37) se puede sustituir por el segmento de $-(1/\tau)$ a $1 - (1/\tau)$. Así,

$$I(N) = \int_{-1/\tau}^{1-1/\tau} T_a^3 e^{-2\pi i a N} d\alpha. \quad (38)$$

Consideremos ahora todas las fracciones propias irreducibles a/q con denominadores no superiores a h , y distinguimos en el segmento $-(1/\tau) \leq \alpha \leq 1 - (1/\tau)$ los intervalos «básicos» correspondientes a estas fracciones

$$\frac{a}{q} - \frac{1}{\tau} \leq \alpha \leq \frac{a}{q} + \frac{1}{\tau}; \quad (39)$$

para N suficientemente grande, estos intervalos, como se puede probar⁶, no tienen puntos comunes. De esta manera, el segmento $-(1/\tau) \leq \alpha \leq 1 - (1/\tau)$ se puede descomponer en intervalos básicos e intervalos «complementarios». Representamos $I(N)$ como suma de dos términos

$$I(N) = I_1(N) + I_2(N), \quad (40)$$

donde $I_1(N)$ denota la suma de las integrales sobre los intervalos básicos e $I_2(N)$ es la suma de las integrales sobre los intervalos complementarios. Como se verá más abajo, cuando el número impar N aumenta indefinidamente, también lo hace $I_1(N)$, y

$$\lim_{N \rightarrow \infty} \frac{I_2(N)}{I_1(N)} = 0. \quad (41)$$

⁶ Si dos de tales intervalos alrededor de los puntos a_1/q_1 y a_2/q_2 se cortan, entonces en un punto común tendríamos la ecuación

$$\frac{a_1}{q_1} + \frac{\theta_1}{\tau} = \frac{a_2}{q_2} + \frac{\theta_2}{\tau}, \text{ donde } |\theta_1| \leq 1, \quad |\theta_2| \leq 1,$$

o bien

$$\frac{a_1 q_2 - a_2 q_1}{q_1 q_2} = \frac{\theta_2 - \theta_1}{\tau}.$$

Pero el valor absoluto del primer miembro de esta última ecuación no es menor que $1/q_1 q_2$, es decir, es mayor que $1/h^2$, y el segundo miembro no es mayor que $2/\tau$, es decir, es menor que $2h/N$. Así, si esta última ecuación fuera cierta, ello implicaría la desigualdad $1/h^2 < 2h/N$, que contradice la elección de h .

En virtud de (40) vemos, pues, que el número de representaciones de un número impar N como suma de tres primos aumenta indefinidamente con N , por lo que, en particular, hemos probado la conjetura de Goldbach para todo número impar N suficientemente grande.

Expresión para la integral sobre los intervalos básicos

Sea a perteneciente a uno de los intervalos básicos; por (39), $a = a/q + z$, donde $1 \leq q \leq h$ y $|z| \leq 1/\tau$. Descomponemos la suma (36)

$$T_a = \sum_{p \leq N} e^{2\pi i a p} = \sum_{p \leq N} e^{2\pi i (a/q + z)p},$$

extendida a todos los primos no superiores a N , en sumas parciales $T_{a,M}$ de la forma

$$T_{a,M} = \sum_{M \leq p < M'} e^{2\pi i (a/q + z)p},$$

donde M' se escoge de tal manera que $e^{2\pi i z p}$ difiera «poco» de $e^{2\pi i z M}$; como solamente intentamos dar la *idea* del método de Vinogradov, y no una demostración del teorema de Goldbach-Vinogradov, no estableceremos con precisión lo que entendemos por la expresión «difiere poco»; en su demostración, Vinogradov trata con desigualdades rigurosamente definidas, que llevan consigo un gran número de cálculos. De este modo

$$T_{a,M} \approx e^{2\pi i M z} \sum_{M \leq p < M'} e^{2\pi i (a/q)p} = e^{2\pi i M z} T_{a/q,M}, \quad (42)$$

donde el símbolo $\approx$ significa que la primera de las tres expresiones de la última relación difiere «poco» de la segunda.

Descomponemos además cada una de las sumas

$$T_{a/q,M} = \sum_{M \leq p < M'} e^{2\pi i (a/q)p} \quad (43)$$

en sumas $T_{a/q,M'l}$, extendidas a todos los primos p_l que verifican la relación $M \leq p_l < M'$ y pertenecen a progresiones aritméticas $qx + l$, donde l toma todos los valores desde 0 a $q-1$ que son primos con q . Pero

$$e^{2\pi i (a/q)p} = e^{2\pi i a x + 2\pi i (a/q)l} = e^{2\pi i (a/q)l},$$

y por tanto

$$T_{a/q,M'l} = e^{2\pi i (a/q)l} \cdot \pi(M, M', l), \quad (44)$$

donde $\pi(M, M', l)$ es el número de primos que verifican la condición $M \leq p < M'$ y forman parte de la progresión aritmética $qx + l$. En el desarrollo de la fórmula (14) para el número $\pi(x)$ de primos no superiores a x , se estableció que $\pi(M, M', l)$, para valores de q «pequeños» en comparación con la diferencia $M' - M$, difiere poco de $1/\phi(q) \int_M^{M'} dx/\ln x$, donde $\phi(q)$ es la *función de Euler*. Esta es una función de la teoría de números (es decir, definida para números naturales q) que representa el número de enteros positivos no superiores a q y que son primos con él. De (44) podemos así deducir

$$T_{a/q, M'l} \approx e^{2\pi i(a/q)l} \cdot \frac{1}{\phi(q)} \int_M^{M'} \frac{dx}{\ln x}. \quad (45)$$

En la expresión del segundo miembro de (45), solamente el primer factor depende de l , es decir, de la elección de la progresión aritmética $qx + l$ (ahora consideramos q fijo). Sumando en l , obtenemos

$$T_{a/q, M} \approx \frac{1}{\phi(q)} \int_M^{M'} \frac{dx}{\ln x} \sum_l e^{2\pi i(a/q)l},$$

y además, de (42),

$$T_{a, M} \approx e^{2\pi i Mz} \cdot \frac{1}{\phi(q)} \int_M^{M'} \frac{dx}{\ln x} \cdot \sum_l e^{2\pi i(a/q)l}, \quad (46)$$

donde

$$e^{2\pi i Mz} \int_M^{M'} \frac{dx}{\ln x} \approx \int_M^{M'} \frac{e^{2\pi i z x}}{\ln x} dx,$$

lo cual nos permite sustituir (46) por la relación

$$T_{a, M} \approx \int_M^{M'} \frac{e^{2\pi i z x}}{\ln x} dx \cdot \frac{1}{\phi(q)} \sum_l e^{2\pi i(a/q)l}. \quad (47)$$

Sumando en M queda

$$T_a \approx \int_2^N \frac{e^{2\pi i z x}}{\ln x} dx \cdot \frac{1}{\phi(q)} \sum_l e^{2\pi i(a/q)l}. \quad (48)$$

La suma

$$\sum_l e^{2\pi i(a/q)l},$$

que aparece en el segundo miembro de (48), extendida a todos los números naturales l no superiores a q y primos con él, se puede expresar como una función de la teoría de números $\mu(q)$ definida de la siguiente manera: $\mu(q) = 0$, si q es divisible por el cuadrado de un entero mayor que uno; $\mu(1) = 1$ y $\mu(q) = (-1)^n$, si $q = p_1 p_2 \cdots p_n$, donde $p_1, p_2, \dots, p_n$ son primos distintos. Por tanto, para a y q primos entre sí

$$\sum_l e^{2\pi i(a/q)l} = \mu(q). \quad (49)$$

Por consiguiente, la ecuación (48) se puede escribir en la forma

$$T_a \approx \frac{\mu(q)}{\phi(q)} \int_2^N \frac{e^{2\pi i z x}}{\ln x} dx.$$

Por ser $\mu^3(q) = \mu(q)$ tenemos

$$T_a^3 \approx \frac{\mu(q)}{[\phi(q)]^3} \left(\int_2^N \frac{e^{2\pi i z x}}{\ln x} dx \right)^3, \quad (50)$$

y de la definición de $(I_1 N)$

$$I_1(N) = \sum_{1 \leq a < h} \sum_a \int_{a/q-1/\tau}^{a/q+1/\tau} T_a^3 e^{-2\pi i a N} da, \quad (51)$$

donde, para un q dado, la suma está extendida a todos los a no negativos menores que q . Como $a = a/q + z$, tenemos, como resultado de (50),

$$I_1(N) \approx \sum_{1 \leq a < h} \frac{\mu(q)}{[\phi(q)]^3} \sum_a e^{-2\pi i(a/q)N} \int_{-1/\tau}^{1/\tau} \left(\int_2^N \frac{e^{2\pi i z x}}{\ln x} dx \right)^3 e^{-2\pi i z N} dz. \quad (52)$$

Introducimos la notación

$$R(N) = \int_{-1/\tau}^{1/\tau} \left(\int_2^N \frac{e^{2\pi i z x}}{\ln x} dx \right)^3 e^{-2\pi i z N} dz. \quad (53)$$

De la relación (52) se deduce que

$$I_1(N) \approx R(N) \sum_{1 \leq a < h} \frac{\mu(q)}{[\phi(q)]^3} \sum_a e^{2\pi i(a/q)N}. \quad (54)$$

Aquí debemos llamar la atención sobre el hecho de que $R(N)$ es una expresión analítica, que puede, por tanto, calcularse aproximadamente; en efecto, se obtiene que

$$R(N) \approx \frac{N^2}{2(\ln N)^3}. \quad (55)$$

La expresión que aparece como factor de $R(N)$ en el segundo miembro de (54) difiere «poco» de la suma de la serie infinita

$$S(N) = \sum_{q=1}^{\infty} \frac{\mu(q)}{[\phi(q)]^3} \sum_a e^{-2\pi i(a/q)N}, \quad (56)$$

por lo que, de (54) y (55), se puede establecer que

$$I_1(N) \approx \frac{N^2}{2(\ln N)^3} S(N), \quad (57)$$

o, más exactamente,

$$I_1(N) = \frac{N^2}{2(\ln N)^3} [S(N) + \gamma_1(N)], \quad (58)$$

donde

$$\lim_{N \rightarrow \infty} \gamma_1(N) = 0. \quad (59)$$

Observemos que la expresión de la teoría de números $S(N)$ se puede escribir en la forma

$$S(N) = C \prod_p \left(1 - \frac{1}{p^2 - 3p + 3} \right), \quad (60)$$

donde C es una constante, la multiplicación está extendida a todos los divisores primos del número N , y, como demuestran los cálculos,

$$S(N) > 0,6. \quad (61)$$

Estimación de la integral sobre los intervalos complementarios

Ahora vamos a tratar de obtener una estimación de la suma I_2 de las integrales sobre los intervalos complementarios. Puesto que

el módulo de la integral no excede de la integral del módulo del integrando, y $|e^{-2\pi i \alpha N}| = 1$ para αN real, tenemos

$$|I_2| < \max |T_\alpha| \cdot \int_{-1/\tau}^{1-1/\tau} |T_\alpha|^2 d\alpha, \quad (62)$$

donde $\max |T_\alpha|$ representa el mayor valor de $|T_\alpha|$ para α perteneciente a los intervalos complementarios —hemos reforzado la desigualdad tomando como factor de $\max |T_\alpha|$ la integral extendida a todo el intervalo — $(1/\tau) \leq \alpha \leq 1 - (1/\tau)$.

Pero el cuadrado del módulo de un número complejo es igual al producto del número por su conjugado, de modo que

$$|T_\alpha|^2 = T_\alpha \cdot \bar{T}_\alpha,$$

donde en virtud de (36) tenemos

$$\bar{T}_\alpha = \sum_{p \leq N} e^{-2\pi i \alpha p},$$

ya que $e^{-2\pi i \alpha p} = \cos 2\pi \alpha p - i \sin 2\pi \alpha p$. Por tanto, la desigualdad (62) se puede volver a escribir en la forma

$$|I_2| < \max |T_\alpha| \cdot \int_{-1/\tau}^{1-1/\tau} \sum_{p \leq N} e^{2\pi i \alpha p} \sum_{p_1 \leq N} e^{-2\pi i \alpha p_1} d\alpha,$$

o en la forma

$$|I_2| < \max |T_\alpha| \cdot \int_{-1/\tau}^{1-1/\tau} \sum_{p \leq N} \sum_{p_1 \leq N} e^{2\pi i \alpha (p - p_1)} d\alpha. \quad (63)$$

Pero la integral de la desigualdad (63) (por lo que se ha dicho al principio de la presente sección) representa el número de soluciones en los primos p, p_1 , no superiores a N , de la ecuación $p - p_1 = 0$, o simplemente el número de primos no superiores a N , es decir, $\pi(N)$. En virtud del resultado (12) de Chebichev, tenemos

$$\pi(N) < B \cdot \frac{N}{\ln N},$$

donde B es una constante. De esta manera

$$|I_2| < B \cdot \frac{N}{\ln N} \cdot \max |T_\alpha|, \quad (64)$$

donde, repetimos, $\max |T_a|$ representa el mayor valor de $|T_a|$ sobre los intervalos complementarios. De (58) y (59) se deduce que para completar la demostración del teorema de Goldbach-Vinogradov deberíamos probar ahora que el orden de $\max |T_a|$ es menor que el de $N/(\ln N)^2$; sin embargo, el establecimiento de este hecho es sumamente difícil y constituye la parte esencial de toda la demostración del teorema.

Todo a perteneciente a un intervalo complementario se puede poner en la forma $a = a/q + z$, donde $h < q \leq \tau$ y $|z| \leq 1/q\tau$. El problema, por tanto, se reduce a la estimación del módulo de la suma trigonométrica

$$T_a = \sum_{p \leq N} e^{2\pi i(a/q + z)p}$$

bajo las condiciones dadas. Vinogradov estableció, en particular, que

$$\lim_{N \rightarrow \infty} \frac{\max |T_a|}{\frac{N}{(\ln N)^3}} = 0; \quad (65)$$

aquí hizo uso de una identidad muy importante que descubrió para la función $\mu(n)$ discutida anteriormente.

Desgraciadamente no es posible dar aquí una demostración de la ecuación (65); remitimos al lector interesado al capítulo 10 del libro de Vinogradov *Métodos de sumas trigonométricas en la teoría de números*.

De (65) y (64), como hemos observado, se deduce que

$$\lim_{N \rightarrow \infty} \frac{I_2(N)}{I_1(N)} = 0.$$

De esta manera, en virtud de (40), (58) y (59) tenemos

$$I(N) = \frac{N^2}{2(\ln N)^3} [S(N) + \gamma(N)], \quad (66)$$

donde

$$\lim_{N \rightarrow \infty} \gamma(N) = 0$$

y $S(N)$ tiene el valor (60), por lo que, de (61), $S(N) > 0,6$. Esto completa la demostración del teorema.

§5. *Descomposición de enteros en suma de dos cuadrados; enteros complejos*

La importancia del estudio de los números primos radica sobre todo en el papel fundamental que juegan en la mayoría de las leyes de la teoría de números: ocurre con frecuencia que cuestiones que a primera vista parecen muy alejadas de la divisibilidad, resultan estar, tras consideraciones más precisas, íntimamente relacionadas con la teoría de números primos. Ilustramos esta afirmación con el siguiente ejemplo.

Uno de los problemas de la teoría de números consiste en hallar aquellos números naturales que se puedan descomponer en suma de los cuadrados de dos enteros (no necesariamente distintos de cero).

La ley de formación de la sucesión de números que son suma de dos cuadrados no es evidente. De 1 a 50, por ejemplo, está compuesta por los números 1, 2, 4, 5, 8, 9, 10, 13, 16, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40, 41, 45, 49, 50, sucesión que parece bastante irregular. El matemático francés del siglo XVII Fermat observó que aquí todo depende de cómo se pueda representar el número como producto de primos, es decir, el problema está intrínsecamente relacionado con la teoría de números primos.

Todos los números primos, salvo $p = 2$, son impares, por lo que su división por 4 da de resto 1 (para un primo de la forma $4n + 1$) ó 3 (para un número primo de la forma $4n + 3$).

El problema de expresar un número dado como la suma de dos cuadrados lo estudiaremos bajo los tres epígrafes siguientes.

1. Un número primo p es suma de dos cuadrados si, y sólo si, $p = 4n + 1$.

La demostración de que un número de la forma $4n + 3$ no se puede expresar como suma de dos cuadrados es casi evidente: la suma de los cuadrados de dos números pares es divisible por 4, la de dos números impares tiene de resto dos cuando se divide por 4, y la suma de cuadrados de un número par y uno impar, al dividirla por 4, da de resto 1.

Demostremos ahora un teorema preliminar: si p es primo, entonces $(p - 1)! + 1$ es divisible por p . Los números no divisibles por p , al dividirlos por p , dan los restos 1, 2, 3, $\dots$, $p - 1$. Tomamos un entero r , $1 \leq r \leq p - 1$, y lo multiplicamos por 1, 2, $\dots$, $p - 1$; si volvemos a dividir los productos así obtenidos por p tenemos, como

se demuestra fácilmente, los mismos restos, aunque en general en distinto orden. En particular, entre ellos estará el número 1, es decir, para cada r podemos encontrar un r_1 tal que $r \cdot r_1 = 1 + kp$. Observemos que $r = r_1$ solamente si $r = 1$ o $r = p - 1$. Porque, si $r^2 = 1 + kp$, entonces $(r + 1) \cdot (r - 1)$ es divisible por p ; pero para los números $1 \leq r \leq p - 1$, esto sólo es posible para $r = 1$ y $r = p - 1$. Hallemos el resto de la división de $(p - 1)! = 1 \cdot 2 \cdot \dots \cdot (p - 1)$ por p . En este producto, para cada factor r , excepto 1 y $p - 1$, aparece un correspondiente r_1 , distinto de r , tal que $r \cdot r_1$ da el resto 1. Por consiguiente, el resto de dividir $(p - 1)!$ por p es el mismo que si solamente figuraran los dos factores 1 y $p - 1$, es decir, da el resto $p - 1$. Por tanto, $(p - 1)! + 1$ es divisible por p .

Sea ahora $p = 4n + 1$. Además, escribimos

$$(p - 1)! + 1 = \left\{ 1 \cdot 2 \cdot \dots \cdot \frac{p-1}{2} \right\} \cdot \left\{ \left(p - \frac{p-1}{2} \right) \cdot \dots \cdot (p-2)(p-1) \right\} + 1.$$

El resto de la segunda expresión entre llaves, al dividirla por p , es $(-1)^{p-1/2} [(p - 1)/2]!$. Pero $(p - 1)/2 = 2n$ es un número par, por lo que en este caso $[(p - 1)/2]!^2 + 1$ es también divisible por p . Denotemos por A el resto de dividir $[(p - 1)/2]!$ por p . Es evidente que $A^2 + 1$ es también divisible por p .

Consideremos la expresión $x - Ay$, en la que x e y recorren independientemente los números $0, 1, \dots, [\sqrt{p}]$; (aquí $[x]$ denota la parte entera de x). Obtenemos así $([\sqrt{p}] + 1)^2 \geq p + 1$ valores numéricos para $x - Ay$, que pueden ser distintos o coincidir en algunos casos. Puesto que los distintos restos de la división por p solamente pueden ser $0, 1, 2, \dots, p - 1$, mientras que aquí tenemos al menos $p + 1$ valores para $x - Ay$, deben existir dos pares distintos (x_1, y_1) y (x_2, y_2) tales que $x_1 - Ay_1$ y $x_2 - Ay_2$ tengan el mismo resto al dividirlos por p ; es decir $(x_1 - x_2) - A(y_1 - y_2)$ es divisible por p . Ponemos $x_0 = x_1 - x_2, y_0 = y_1 - y_2$. Evidentemente, $|x_0| < \sqrt{p}, |y_0| < \sqrt{p}$. Como $A^2 + 1$ es divisible por p , se deduce que $y_0^2(A^2 + 1) = (Ay_0)^2 + y_0^2$ también lo es; pero como $x_0 - Ay_0$ es divisible por p , el número $x_0^2 - (Ay_0)^2 = (x_0 - Ay_0)(x_0 + Ay_0)$ lo sigue siendo. Por tanto la cantidad $x_0^2 + y_0^2$, que es igual a $(x_0^2 - (Ay_0)^2 + (Ay_0)^2 + y_0^2)$, es divisible por p . Pero $|x_0| < \sqrt{p}, |y_0| < \sqrt{p}$. De aquí que $x_0^2 + y_0^2 = 0$ o $x_0^2 + y_0^2 = p$. Lo primero es imposible, ya que los pares (x_1, y_1) y (x_2, y_2) son distintos. Por tanto, un número primo de la forma $4n + 1$ es representable como suma de dos cuadrados.

2. Volvemos a la descomposición de un entero arbitrario en suma de dos cuadrados. Es fácil establecer la identidad

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2.$$

Esta identidad prueba que el producto de dos enteros que son suma de dos cuadrados es a su vez suma de dos cuadrados. De aquí que el producto de potencias cualesquiera de números primos de la forma $4n + 1$ (o iguales a 2) es la suma de dos cuadrados. Puesto que al multiplicar la suma de dos cuadrados por un cuadrado vuelve a dar la suma de dos cuadrados, cualquier número en el que los factores primos de la forma $4n + 3$ aparezcan con potencias pares es suma de dos cuadrados.

3. Demostremos ahora que, si un número primo de la forma $4n + 3$ figura en un número con potencia impar, este no se puede representar como suma de dos cuadrados. La cuestión original quedará entonces completamente resuelta.

Consideraremos números complejos de la forma $a + bi$, donde a y b son enteros ordinarios. Tales números complejos se denominarán *enteros complejos*. Si un entero N es suma de dos cuadrados $N = a^2 + b^2$, entonces $N = (a + bi)(a - bi) = a \cdot \bar{a}$ (donde $\bar{a}$ denota el conjugado del número a), es decir, N se factoriza, dentro del dominio de los enteros complejos, en factores complejos conjugados.

En este dominio podemos construir una teoría de la divisibilidad completamente análoga a la teoría de la divisibilidad en el dominio de los enteros ordinarios. Diremos que el entero complejo a es divisible por el entero complejo β , si a/β es a su vez un entero complejo. Existen solamente cuatro enteros complejos que dividen a 1, a saber, 1, -1 , i y $-i$. Diremos que un entero complejo a es primo si no tiene otros divisores que 1, -1 , i , $-i$, a , $-a$, ai , $-ai$. Pero ahora el problema resuelto en la primera parte tendrá un sentido diferente; ahora los números de la forma $4n + 1$ (o iguales a 2) que antes eran primos dejarán de serlo en el dominio de enteros complejos, mientras que es fácil probar que los primos de la forma $4n + 3$ siguen siéndolo.

En efecto, si $p = a\beta$, entonces $p = \bar{a}\bar{\beta}$ y $p^2 = a\bar{a}\beta\bar{\beta}$. Pero $a\bar{a}$ y $\beta\bar{\beta}$ son enteros positivos ordinarios; y $p \neq a\bar{a}$, ya que los números primos de la forma $4n + 3$ no son suma de dos cuadrados. Esto significa que $a\bar{a} = 1$; por tanto a puede ser solamente ± 1 o $\pm i$, de modo que p no tiene otros divisores que los triviales.

Para enteros complejos el teorema de la descomposición única en factores primos sigue siendo válido. La unicidad significa aquí,

naturalmente, que se hace caso omiso del orden de los factores y también de aquellos de la forma $1, -1, i, -i$.

Sea N suma de dos cuadrados, $N = a\bar{a}$. Sea p un número primo de la forma $4n + 3$. Calculemos qué potencia de p aparece en el número N . Dado que p sigue siendo primo en el dominio complejo, basta calcular qué potencia de p aparece en a y en $\bar{a}$. Pero estas potencias son iguales, de modo que p figura necesariamente en N con potencia par, lo cual prueba la proposición.

El descubrimiento de que es posible construir una fecunda teoría de divisibilidad en otros campos distintos del de los números racionales amplió enormemente el horizonte de los matemáticos del siglo XIX. El desarrollo de estas ideas exigió la creación de nuevos conceptos generales de matemáticas, tales como, por ejemplo, anillos e ideales. La importancia de estos conceptos en el momento actual ha superado con creces el marco de la teoría de números.

§1. *Las leyes de la probabilidad*

Las leyes más simples de las ciencias naturales son aquellas que establecen las condiciones bajo las cuales algún suceso de interés para nosotros ocurrirá, o no ocurrirá, pero en ambos casos con certeza; estas condiciones se pueden expresar en una de las dos formas siguientes:

1. Si un sistema (un conjunto o colección) de condiciones S se realiza, entonces el suceso A ocurre con certeza;

2. Si un sistema de condiciones S se realiza, entonces el suceso A no puede ocurrir.

En el primer caso, el suceso A , con respecto al sistema de condiciones S , se dice que es un suceso «cierto» o «necesario», y en el segundo, un suceso «imposible». Por ejemplo, a la presión atmosférica y a una temperatura t entre 0° y 100° (el sistema de condiciones S) el agua necesariamente está en estado líquido (el suceso A_1 es cierto) y no puede estar en estado sólido o gaseoso (los sucesos A_2 y A_3 son imposibles).

Un suceso A , que bajo un sistema de condiciones S a veces ocurre y a veces no, se denomina aleatorio con respecto al sistema de condiciones. Esto suscita la cuestión: ¿demuestra la casualidad del

suceso A la ausencia de una ley que relacione el sistema de condiciones S con A ? Por ejemplo, supongamos que las bombillas de un tipo específico, producidas en cierta fábrica (condición S), a veces lucen durante más de 2.000 horas (suceso A), pero a veces se funden y quedan inservibles antes de transcurrir este tiempo. ¿Es posible que los resultados de los experimentos para ver si una determinada bombilla luce o no durante 2.000 horas sirvan para valorar la producción de la fábrica? ¿O debemos limitarnos a indicar el período (digamos 500 horas) durante el cual prácticamente todas las bombillas lucen sin fallos, y el período (digamos 10.000 horas) a partir del cual prácticamente todas quedan inservibles? Está claro que describir la vida de una bombilla mediante una desigualdad de la forma $500 \leq T \leq 10.000$ es de escasa utilidad para el consumidor. Recibirá una información mucho más valiosa si le decimos que aproximadamente en el 80 por 100 de los casos las bombillas lucen no

menos de 2.000 horas. Una valoración más completa aún de la calidad de las bombillas consiste en indicar para todo T el porcentaje $\nu(T)$ de las que duran no menos de T horas, por ejemplo como en la gráfica de la figura 1.

En la práctica la curva $\nu(T)$ se halla verificando una muestra suficiente-

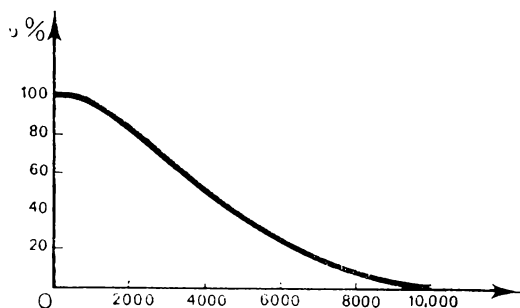


FIG. 1

mente grande (100-200) de lámparas. Naturalmente, la curva hallada de esta manera tiene verdadero interés solamente cuando represente fielmente una ley real que gobierne no sólo la muestra dada, sino todas las bombillas fabricadas con un material de una determinada calidad y bajo condiciones tecnológicas conocidas; es decir, solamente si al llevar a cabo los mismos experimentos con otra muestra se obtienen aproximadamente los mismos resultados (es decir, la nueva curva $\nu(T)$ se diferencia poco de la que se obtuvo en la primera muestra). En otras palabras, la ley estadística expresada por las curvas $\nu(T)$ para las distintas muestras es sólo un reflejo de la ley probabilística que relaciona la vida útil de

una bombilla con los materiales y las condiciones tecnológicas de su fabricación.

Esta ley de probabilidad está dada por una función $P(T)$, donde $P(T)$ es la probabilidad de que una bombilla (fabricada bajo condiciones dadas) luzca no menos de T horas.

La afirmación de que un suceso A ocurre en las condiciones S con una determinada probabilidad

$$P(A/S) = p$$

equivale a decir que en una serie suficientemente larga de ensayos (es decir, realizaciones del sistema de condiciones S), las frecuencias

$$\nu_r = \frac{\mu_r}{n_r}$$

de ocurrencia del suceso A (donde n_r es el número de ensayos realizados en la r -ésima serie, y μ_r el número de ellos en que ocurre A) son aproximadamente idénticas unas a otras y están próximas a p .

La hipótesis de la existencia de una constante $p = P(A/S)$ (determinada objetivamente por la relación entre el sistema de condiciones S y el suceso A) tal que las frecuencias ν se aproximen más (hablando en términos generales) a p a medida que aumenta el número de ensayos, está justificada en la práctica para una amplia clase de sucesos. Los sucesos de este tipo se denominan normalmente *aleatorios* o *estocásticos*.

Este ejemplo pertenece a las leyes de probabilidad para la producción en serie. La realidad de tales leyes no se puede poner en duda, y constituyen la base de importantes aplicaciones prácticas en el control estadístico de la calidad. De tipo análogo son las leyes de probabilidad para la dispersión de proyectiles, fundamentales en la teoría del tiro con armas de fuego. Ya que históricamente es una de las aplicaciones más antiguas de la teoría de probabilidades a problemas de la técnica, volveremos más adelante sobre algunos problemas sencillos de dicha teoría.

Lo que hemos dicho acerca de la «proximidad» de la frecuencia ν a la probabilidad p para un gran número n de pruebas es muy vago; no hemos dicho nada respecto al valor de la diferencia $\nu - p$ para cualquier n . El grado de proximidad de ν a p está calculado en la sección 3. Es interesante observar que en este caso resulta inevitable

cierta indeterminación. La propia afirmación de que ν y p están próximas una a otra tiene solamente carácter probabilístico, como se pone en evidencia cuando tratamos de precisar la situación.

§2. *Los axiomas y fórmulas básicas de la teoría elemental de probabilidades*

Puesto que no hay duda de que las leyes estadísticas son de gran importancia, examinaremos los métodos para su estudio. En primer lugar uno piensa en la posibilidad de proceder de una manera puramente empírica. Ya que las leyes probabilísticas sólo aparecen en procesos masivos, es natural imaginar que para descubrir cualquier ley debemos realizar un experimento en el que intervengan grandes números.

Tal idea, sin embargo, es sólo correcta en parte. Tan pronto como hayamos establecido ciertas leyes de probabilidad mediante un experimento, podemos proceder a deducir otras nuevas por medios lógicos o de cálculo, bajo ciertas hipótesis generales. Antes de indicar cómo se hace esto, debemos enumerar algunas definiciones y fórmulas básicas de la teoría de probabilidades.

De la representación de la probabilidad como el valor standard de la frecuencia $\nu = m/n$, donde $0 \leq m \leq n$, y por tanto $0 \leq \nu \leq 1$, se deduce que la probabilidad $P(A)$ de un suceso A se debe suponer comprendida entre cero y uno¹

$$0 \leq P(A) \leq 1. \quad (1)$$

Dos sucesos se dice que son mutuamente excluyentes si no pueden ocurrir ambos a la vez (bajo el sistema de condiciones S). Por ejemplo, al lanzar un dado, que salga un número par y un tres son sucesos mutuamente excluyentes. Un suceso A se dice *unión* de dos sucesos A_1 y A_2 si consiste en la ocurrencia de al menos uno de los sucesos A_1 , A_2 . Por ejemplo, al lanzar un dado, el suceso A , consistente en que salga 1, 2 ó 3, es la unión de A_1 y A_2 , donde A_1 es que salga 1 ó 2, y A_2 , que salga 2 ó 3. Es fácil ver que para el número de repeticiones m_1 , m_2 y m de dos sucesos mutuamente excluyentes A_1 , A_2 y su unión $A = A_1 \cup A_2$, tenemos la ecuación $m = m_1 + m_2$, o para las frecuencias correspondientes $\nu = \nu_1 + \nu_2$.

¹ Por brevedad cambiamos ahora $P(A/S)$ por $P(A)$.

Esto conduce, naturalmente, al siguiente axioma para la adición de probabilidades:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2), \quad (2)$$

si los sucesos A_1 y A_2 son mutuamente excluyentes y $A_1 \cup A_2$ denota su unión.

Además, para un suceso U cuya ocurrencia sea segura, tomamos naturalmente

$$P(U) = 1. \quad (3)$$

Toda la teoría matemática de probabilidades está construida sobre la base de axiomas sencillos del tipo (1), (2) y (3). Desde el punto de vista de la matemática pura, la *probabilidad* es una función numérica de los «sucesos», con ciertas propiedades determinadas por los axiomas. Las propiedades de la probabilidad, expresadas por las fórmulas (1), (2) y (3), sirven como base suficiente para la construcción de lo que se denomina teoría elemental de probabilidades, en cuya axiomatización no entran los conceptos de suceso, unión de sucesos e intersección de sucesos, que definiremos después. Para el principiante es más útil limitarse a un conocimiento intuitivo de los términos «suceso» y «probabilidad», pero dándose cuenta de que, si bien el sentido de estos términos no se puede formalizar completamente en la vida práctica, ello no afecta a la completa precisión formal de una presentación axiomatizada, puramente matemática, de la teoría de probabilidades.

La unión de cualquier número de sucesos $A_1, A_2, \dots, A_s$ se define como aquel suceso A consistente en la ocurrencia de al menos uno de los A_k . Del axioma de adición obtenemos fácilmente, para un número finito de sucesos mutuamente excluyentes dos a dos $A_1, A_2, \dots, A_s$ y su unión A , que

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_s)$$

(teorema de adición de probabilidades).

Si la unión de estos sucesos es un suceso cierto (es decir, si bajo el sistema de condiciones S es seguro que uno de los A_k ocurre), entonces

$$P(A_1) + P(A_2) + \dots + P(A_s) = 1.$$

En este caso la familia $A_1, \dots, A_s$ se denomina *sistema completo* de sucesos.

Consideremos ahora dos sucesos A y B que, en general, no serán mutuamente excluyentes. El suceso C es la intersección de A y B , escrito $C = AB$, si C consiste en la ocurrencia tanto de A como de B ². Por ejemplo, si A consiste en obtener un número par al lanzar un dado y B en obtener un múltiplo de tres, entonces C consiste en sacar un seis.

Supongamos que, en un gran número de ensayos n , el suceso A se repite m veces y el B l veces, en k de las cuales B ocurre junto con A . El cociente k/m se denomina frecuencia condicionada del suceso B bajo la condición A . Las frecuencias k/m , m/n y k/n están relacionadas por la fórmula

$$\frac{k}{m} = \frac{k}{n} : \frac{m}{n},$$

que naturalmente da lugar a la siguiente definición:

La probabilidad condicionada $P(B/A)$ del suceso B bajo la condición A es el cociente

$$P(B/A) = \frac{P(AB)}{P(A)}.$$

Aquí se supone, naturalmente, que $P(A) \neq 0$.

Si los sucesos A y B no están esencialmente relacionados uno con otro, entonces es natural suponer que B no aparecerá ni más frecuentemente ni menos al ocurrir A , es decir, que aproximadamente $k/m \sim l/n$, o bien

$$\frac{k}{n} = \frac{k}{m} \frac{m}{n} \sim \frac{l}{n} \frac{m}{n}.$$

En esta última ecuación aproximada $m/n = v_A$ es la frecuencia del suceso A , $l/n = v_B$ es la frecuencia de B y, finalmente, $k/n = v_{AB}$ es la frecuencia de la intersección de A y B .

Vemos que estas frecuencias están relacionadas por

$$v_{AB} \sim v_A v_B.$$

Para las probabilidades de los sucesos A , B y AB , es por consiguiente natural aceptar la ecuación exacta correspondiente

$$P(AB) = P(A) \cdot P(B). \quad (4)$$

² Análogamente, la intersección C de varios sucesos $A_1, A_2, \dots, A_s$ consiste en que ocurran todos ellos.

La ecuación (4) sirve para definir la *independencia* de dos sucesos A y B .

Análogamente podemos definir la independencia de cualquier número de sucesos; o bien dar una definición de la independencia de cualquier número de experimentos, lo cual significa, en términos generales, que el resultado de una parte cualquiera de los experimentos no depende del resultado de los restantes³.

Calculemos ahora la probabilidad P_k de k repeticiones de un cierto suceso A en n ensayos independientes, en cada uno de los cuales la probabilidad p de que ocurra A es la misma. Denotemos por $\bar{A}$ el suceso que consiste en que no ocurra A . Es evidente que

$$P(\bar{A}) = 1 - P(A) = 1 - p.$$

Por la definición de la independencia de experimentos es fácil ver que la probabilidad de una sucesión específica formada por k repeticiones de A y $n - k$ no-apariciones de A es igual a

$$p^k(1 - p)^{n-k}. \quad (5)$$

Así, por ejemplo, para $n = 5$ y $k = 2$ la probabilidad de obtener la sucesión $A\bar{A}A\bar{A}A$ será $p(1 - p)p(1 - p)(1 - p) = p^2(1 - p)^3$.

Por el teorema de adición de probabilidades, p_k será igual a la suma de las probabilidades de todas las sucesiones con k apariciones y $n - k$ no-apariciones del suceso A , es decir, p_k será igual, en virtud de (5), al producto del número de tales sucesiones por $p^k(1 - p)^{n-k}$. Evidentemente, habrá tantas sucesiones como combinaciones de n elementos tomados de k en k , ya que los k resultados positivos pueden ocupar k lugares cualesquiera en la sucesión de n ensayos.

Finalmente obtenemos

$$P_k = C_n^k p^k (1 - p)^{n-k} \quad (k = 0, 1, 2, \dots, n) \quad (6)$$

(que se denomina distribución binomial).

³ Un significado más exacto de *experimentos independientes* es el siguiente. Dividimos los n experimentos de cualquier forma en dos grupos; el suceso A consiste en que todos los experimentos del primer grupo tengan ciertos resultados prefijados, y el suceso B , en que los experimentos del segundo grupo tengan también ciertos resultados prefijados. Los experimentos se denominan independientes (como colección) si, para descomposiciones arbitrarias en dos grupos y resultados prefijados arbitrariamente, los sucesos A y B son independientes en el sentido de (4).

En §4 volveremos a considerar el sentido objetivo de la independencia de sucesos en el mundo real.

Para ver cómo se aplican las definiciones y fórmulas, consideremos un ejemplo que surge en la teoría de armas de fuego.

Supongamos que cinco impactos son suficientes para la destrucción del blanco. Lo que nos interesa es la cuestión de si podemos suponer que 40 disparos aseguran los cinco impactos necesarios. Una resolución puramente empírica del problema procedería como sigue. Para dimensiones del blanco y alcance dados, efectuamos un gran número (digamos 200) de ensayos, cada uno consistente en 40 disparos, y determinamos cuántas de estas pruebas producen al menos cinco impactos. Si este resultado se logra, por ejemplo, con 195 de las 200, entonces la probabilidad P es aproximadamente igual a

$$P = \frac{195}{200} = 0,975.$$

Si procedemos de esta forma puramente empírica, gastaremos 8.000 proyectiles para resolver un problema particular muy sencillo. En la práctica, naturalmente, no se hace así. En lugar de ello, comenzamos la investigación suponiendo que la dispersión de los proyectiles para un alcance dado es independiente de las dimensiones del blanco. Ocurre que las desviaciones longitudinales y laterales a partir del punto medio de caída de los proyectiles siguen una ley respecto a la frecuencia de desviaciones que se ilustra en la figura 2. La letra B denota la llamada desviación probable. La desviación probable, en

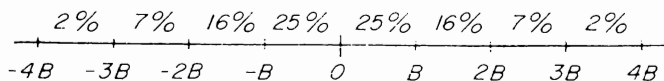


FIG. 2.

general, es distinta para las desviaciones longitudinal y lateral y aumenta con la distancia. Las desviaciones probables para diferentes distancias de cada tipo de cañón y de proyectil se hallan empíricamente mediante prácticas de tiro en un campo de artillería. Pero la subsiguiente resolución de todos los posibles problemas especiales del tipo descrito se lleva a cabo solamente mediante cálculos.

Por sencillez, suponemos que el blanco es rectangular; uno de los lados tiene la dirección de la línea de tiro y una longitud de dos desviaciones longitudinales probables, mientras que el otro es per-

pendicular a la línea de tiro y su longitud es igual a dos desviaciones laterales probables. Suponemos además que el alcance está bien establecido, por lo que la trayectoria media de los proyectiles pasa por el centro del blanco (Fig. 3).

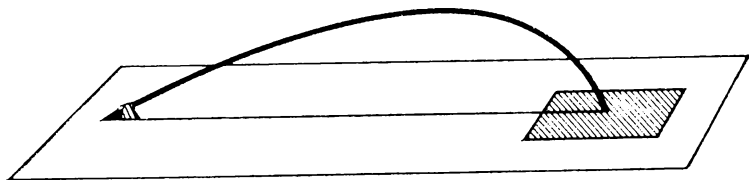


FIG. 3

Suponemos también que las desviaciones laterales y longitudinales son independientes⁴. Entonces, para que un proyectil dado caiga sobre el blanco, es necesario y suficiente que sus desviaciones longitudinal y lateral no excedan de las correspondientes desviaciones probables. Según vemos en la figura 2, cada uno de estos sucesos se observará para el 50 por 100 de los proyectiles disparados, es decir, con probabilidad $1/2$. La intersección de los dos sucesos ocurrirá para el 25 por 100 aproximadamente de los proyectiles disparados; es decir, la probabilidad de que un proyectil determinado dé en el blanco será igual a

$$p = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

y la probabilidad de errar el tiro con un proyectil dado será

$$q = 1 - p = 1 - \frac{1}{4} = \frac{3}{4},$$

Suponiendo que los impactos representan sucesos independientes, y aplicando la fórmula binomial (6), hallamos que la probabilidad de obtener exactamente k impactos en 40 disparos es

$$P_k = C_{40}^k p^k q^{40-k} = \frac{40 \cdot 39 \cdots (39 - k)}{1 \cdot 2 \cdots k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{40-k}.$$

⁴ La hipótesis de independencia viene confirmada por la experiencia.

Lo que nos interesa es la probabilidad de obtener no menos de cinco impactos, y ésta viene ahora expresada mediante la fórmula

$$P = \sum_{k=5}^{40} P_k.$$

Sin embargo es más sencillo calcular esta probabilidad mediante la fórmula $P = 1 - Q$, donde

$$Q = \sum_{k=0}^4 P_k$$

es la probabilidad de obtener menos de cinco impactos.

Tenemos que

$$P_0 = \left(\frac{3}{4}\right)^{40} \sim 0,00001,$$

$$P_1 = 40 \left(\frac{3}{4}\right)^{39} \frac{1}{4} \sim 0,00013,$$

$$P_2 = \frac{40 \cdot 39}{2} \left(\frac{3}{4}\right)^{38} \left(\frac{1}{4}\right)^2 \sim 0,00087,$$

$$P_3 = \frac{40 \cdot 39 \cdot 38}{2 \cdot 3} \left(\frac{3}{4}\right)^{37} \left(\frac{1}{4}\right)^3 \sim 0,0037,$$

$$P_4 = \frac{40 \cdot 39 \cdot 38 \cdot 37}{2 \cdot 3 \cdot 4} \left(\frac{3}{4}\right)^{36} \left(\frac{1}{4}\right)^4 \sim 0,0113,$$

por lo que

$$Q = 0,016, \quad P = 0,984.$$

La probabilidad P así obtenida está algo más próxima a la certeza que lo que suele aceptarse como suficiente en la teoría de armas de fuego. Lo más corriente es determinar el número de proyectiles necesarios para garantizar el resultado con una probabilidad de 0,95.

El ejemplo anterior está algo esquematizado, pero muestra con suficiente detalle la importancia práctica del cálculo de probabilidades. Después de establecer empíricamente la dependencia de las des-

viaciones probables respecto de la distancia (para lo cual no es necesario disparar muchos proyectiles), hemos podido obtener, con cálculos sencillos, las respuestas a cuestiones de la más diversa índole. La situación es la misma en cualquier otro campo en que la influencia colectiva de muchos factores aleatorios conduzca a una ley estadística. El examen directo del cúmulo de observaciones pone en claro solamente las leyes estadísticas más simples; descubre sólo algunas de las probabilidades básicas en juego. Pero entonces, mediante las leyes de la teoría de probabilidades, utilizamos estas probabilidades más simples para calcular las de sucesos más complicados y deducir las leyes estadísticas que los gobiernan.

A veces es posible evitar completamente un cúmulo excesivo de material estadístico, cuando las probabilidades se pueden determinar mediante consideraciones suficientemente convincentes de simetría. Por ejemplo, a la conclusión de que un dado (un cubo hecho de un material homogéneo) caerá con igual probabilidad sobre cada una de sus caras al lanzarlo a suficiente altura, se llegó mucho antes de que existiera una acumulación sistemática de datos para comprobarlo mediante la observación. En los tres últimos siglos, se han llevado a cabo experimentos sistemáticos de esta índole, sobre todo por autores de libros de texto sobre la teoría de probabilidades, cuando la teoría de probabilidades era ya una ciencia bien desarrollada. Los resultados de estos experimentos fueron satisfactorios, pero la cuestión de extenderlos a casos análogos apenas despertó interés. Por ejemplo, por lo que sabemos, nadie ha realizado experimentos suficientemente extensos con dados homogéneos de doce caras. Pero no cabe duda de que, si se hubieran hecho 12.000 lanzamientos, el dado de doce lados habría caído sobre cada una de las caras en mil casos, aproximadamente.

Las probabilidades básicas deducidas de la simetría u homogeneidad también juegan un gran papel en muchos problemas científicos serios, por ejemplo, en los de colisiones o aproximación de moléculas en el movimiento aleatorio de un gas; otro caso en que el éxito ha sido igualmente grande es el del movimiento de las estrellas en una galaxia. Naturalmente, en estos casos más delicados preferimos comprobar nuestras suposiciones teóricas comparándolas con la observación o la experiencia.

§3. La ley de los grandes números y teoremas del límite

Es completamente natural desear dar una mayor precisión cuantitativa a la proposición de que en una larga serie de pruebas la frecuencia de un suceso se «aproxima» a su probabilidad. Pero antes conviene formarnos una idea clara de la delicada naturaleza del problema. En los casos más típicos de la teoría de probabilidades, la situación es tal que en una larga serie arbitraria de pruebas es posible obtener uno de los dos valores extremos para la frecuencia

$$\frac{\mu}{n} = \frac{n}{n} = 1 \quad \text{y} \quad \frac{\mu}{n} = \frac{0}{n} = 0.$$

Así, cualquiera que sea el número de pruebas n , es imposible asegurar con absoluta certeza que tendremos, por ejemplo, la desigualdad

$$\left| \frac{\mu}{n} - p \right| < \frac{1}{10}.$$

Por ejemplo, si el suceso A es sacar un seis con un dado, entonces en n tiradas la probabilidad de obtener un seis en cada una de ellas es $(1/6)^n > 0$; dicho con otras palabras, con probabilidad $(1/6)^n$ tendremos una frecuencia de sacar un seis igual a *uno*; y con probabilidad $(1 - 1/6)^n > 0$ no saldrá un seis en ninguna, es decir, la frecuencia de sacar un seis será igual a *cero*.

En ningún problema similar se puede hacer una estimación no trivial de la proximidad de la frecuencia a la probabilidad con absoluta certeza, sino solamente con probabilidad menor que uno. Por ejemplo, se puede probar que en ensayos independientes⁵, con probabilidad constante p de que un suceso dado ocurra en cada ensayo, la desigualdad

$$\left| \frac{\mu}{n} - p \right| < 0,02 \tag{7}$$

para la frecuencia μ/n se verificará, para $n = 10.000$ (y cualquier p), con probabilidad

$$P > 0,9999. \tag{8}$$

⁵ La demostración de la estimación (8) se discute más adelante en esta sección.

Ante todo hay que señalar que en esta formulación la estimación cuantitativa de la proximidad de la frecuencia μ/n a la probabilidad p implica la introducción de una nueva probabilidad P .

El sentido práctico de la estimación (8) es éste: si realizamos N conjuntos de n pruebas cada uno, y encontramos los M conjuntos en los que se verifica la desigualdad (7), entonces para N suficientemente grande tendríamos aproximadamente

$$\frac{M}{N} \approx P > 0,9999. \quad (9)$$

Pero, si deseamos determinar la relación (9) con más precisión, bien respecto al grado de proximidad de M/N a P , bien respecto a la seguridad con que podamos afirmar que (9) se verifica, debemos recurrir a consideraciones generales del tipo tratado anteriormente al discutir lo que entendíamos por proximidad de μ/n a p . Tales consideraciones podemos repetirlas tantas veces como queramos, pero es evidente que este procedimiento nunca nos eximirá de la necesidad, en la última etapa, de referirnos a probabilidades en el impreciso sentido primitivo de este término.

Sería completamente erróneo pensar que las dificultades de esta clase son peculiares en algún modo de la teoría de probabilidades. En una investigación matemática de sucesos reales, siempre construimos un modelo de ellos. Las discrepancias entre el desarrollo real de los sucesos y el modelo teórico pueden, a su vez, convertirse en objeto de investigación matemática. Pero para estas discrepancias debemos construir otro modelo que utilizaríamos ya sin un análisis matemático formal de las discrepancias que volverían a surgir en la práctica.

Observemos, por otra parte, que en una aplicación real de la estimación ⁶

$$P \left\{ \left| \frac{\mu}{n} - p \right| < 0,02 \right\} > 0,9999 \quad (10)$$

a una serie de n pruebas estamos recurriendo ya a ciertas consideraciones de simetría: la desigualdad (10) indica que para un número muy grande N de series de pruebas, la relación (7) se verificará en no menos del 99,99 por 100 de los casos; entonces es natural esperar con gran confianza que la desigualdad (7) se aplique en particular

⁶ Esta es la notación aceptada para la estimación (8) de la probabilidad de la desigualdad (7).

a aquella sucesión de n pruebas que nos interese, aunque solamente podemos esperar esto si tenemos razones para suponer que la situación de esta sucesión entre las demás es normal, es decir, que no tiene características especiales.

Las probabilidades que podemos decidarnos a despreciar difieren de una situación práctica a otra. Observamos anteriormente que nuestros cálculos preliminares para el gasto de proyectiles necesarios para producir un resultado dado cumplen la norma de que el problema ha de resolverse con probabilidad 0,95, es decir, que las probabilidades despreciadas no excedan de 0,05. La norma se explica por el hecho de que si hubiéramos hecho los cálculos despreciando una probabilidad de sólo 0,01, por ejemplo, habríamos necesitado un gasto mucho mayor de proyectiles, por lo que en la práctica llegaríamos a la conclusión de que no se podría realizar el trabajo en el tiempo disponible, o con la provisión dada de proyectiles.

En investigaciones científicas debemos restringirnos a veces a métodos estadísticos calculados sobre la base de despreciar probabilidades de 0,05, aunque esta práctica sólo debería adoptarse en aquellos casos en que la acumulación de datos más extensos sea muy difícil. Como ejemplo de tal método consideremos el siguiente problema. Suponemos que bajo condiciones específicas el medicamento usual para tratar cierta enfermedad produce resultados positivos en el 50 por 100 de los casos, es decir, con probabilidad 0,5. Se propone un nuevo preparado, y para comprobar sus ventajas planeamos utilizarlo en diez casos, elegidos sin parcialidad entre los pacientes aquejados de la enfermedad. Convenimos entonces en que las ventajas del nuevo producto se considerarán comprobadas si da un resultado positivo en no menos de ocho casos de los diez. Es fácil calcular que tal proceso lleva consigo el desprecio de probabilidades del orden de 0,05 de obtener un resultado erróneo, es decir, de indicar una ventaja del nuevo producto cuando en realidad es igual de eficaz o incluso peor que el antiguo. Porque, si en cada uno de los diez experimentos la probabilidad de un resultado positivo es igual a p , entonces la probabilidad de obtener en diez experimentos 10, 9 u 8 resultados positivos, es respectivamente igual a

$$P_{10} = p^{10}, \quad P_9 = 10p^9(1 - p), \quad P_8 = 45p^8(1 - p)^2.$$

Para el caso $p = 1/2$ la suma de éstos es

$$P = P_{10} + P_9 + P_8 = \frac{56}{1024} \sim 0,05.$$

De esta forma, bajo la hipótesis de que el nuevo preparado es, en efecto, igual de efectivo que el antiguo, nos arriesgamos con una probabilidad del orden de 0,05 a equivocarnos al declarar que el nuevo producto es mejor que el antiguo. Para reducir esta probabilidad hasta un 0,01, sin aumentar el número de experimentos $n = 10$, necesitaremos convenir que la ventaja del nuevo producto estará demostrada si da un resultado positivo en nueve casos de los diez, como mínimo. Si esta exigencia les parece demasiado fuerte a los partidarios del nuevo compuesto, sería necesario hacer un número de experimentos mucho mayor que 10. Por ejemplo, para $n = 100$, si convenimos en que la ventaja del nuevo preparado está probada para $\mu > 65$, entonces la probabilidad de error será solamente $P \approx 0,0015$.

Para investigaciones científicas serias, una norma de 0,05 es claramente insuficiente; pero incluso en materias tan académicas y circunstanciales como el tratamiento de observaciones astronómicas, se acostumbra a despreciar probabilidades de error de 0,001 ó 0,003. Por otra parte, algunos de los resultados científicos basados en las leyes de probabilidad son considerablemente más seguros, incluso, que esto; es decir, llevan consigo el desechar probabilidades más pequeñas. Volveremos sobre esta cuestión más adelante.

En los anteriores ejemplos, hemos utilizado casos particulares de la fórmula binomial (6)

$$P_m = C_n^m p^m (1-p)^{n-m}$$

para la probabilidad de obtener exactamente m resultados positivos en n pruebas independientes, en cada una de las cuales un resultado positivo tiene probabilidad p . Consideremos, mediante esta fórmula, la cuestión planteada al principio de esta sección concerniente a la probabilidad

$$P = \mathbf{P} \left\{ \left| \frac{\mu}{n} - p \right| < \epsilon \right\}, \quad (11)$$

donde μ es el número de resultados positivos⁷. Evidentemente, esta

⁷ Aquí μ toma los valores $m = 0, 1, \dots, n$, con probabilidad P_m ; por ejemplo,

$$\mathbf{P}(\mu = m) = P_m.$$

probabilidad se puede escribir como la suma de aquellas P_m para las cuales m verifica la desigualdad

$$\left| \frac{m}{n} - p \right| < \epsilon, \quad (12)$$

es decir, en la forma

$$P = \sum_{m=m_1}^{m_2} P_m, \quad (13)$$

donde m_1 es el menor de los valores de m que verifican la desigualdad (12), y m_2 , el mayor.

La fórmula (13) para un n razonablemente grande es poco conveniente para el cálculo directo, hecho que explica la gran importancia de la fórmula asintótica descubierta por de Moivre para $p = 1/2$ y por Laplace para un p cualquiera. Esta fórmula nos permite hallar P_m de una manera muy sencilla y estudiar su comportamiento para n grande. La fórmula en cuestión es

$$P_m \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(m-np)^2/2np(1-p)}. \quad (14)$$

Si p no está demasiado próximo a cero o a uno, es suficientemente exacta incluso para un n del orden de 100. Si ponemos

$$t = \frac{m - np}{\sqrt{np(1-p)}}, \quad (15)$$

entonces (14) se convierte en

$$P_m \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-t^2/2}. \quad (16)$$

De (13) y (16) se puede deducir una representación aproximada de la probabilidad (11)

$$P \sim \frac{1}{\sqrt{2\pi}} \int_{-T}^T e^{-t^2/2} dt = F(T), \quad (17)$$

donde

$$T = \epsilon \sqrt{\frac{n}{p(1-p)}}. \quad (18)$$

La diferencia entre los dos miembros de (17) para p fijo, distinto de cero y uno, tiende a cero uniformemente, con respecto a ϵ , cuando $n \rightarrow \infty$. Para la función $F(T)$ se han construido tablas detalladas. He aquí un pequeño fragmento de ellas

T	1	2	3	4
F	0,68269	0,95450	0,99730	0,99993

Para $T \rightarrow \infty$ los valores de la función $F(T)$ convergen a uno.

De la fórmula (17) deducimos una estimación de la probabilidad

$$P \left\{ \left| \frac{\mu}{n} - p \right| < 0,02 \right\}$$

para $n = 10.000$. Puesto que

$$T = \frac{2}{\sqrt{p(1-p)}},$$

tenemos

$$P \approx F\left(\frac{2}{\sqrt{p(1-p)}}\right).$$

Como la función $F(T)$ es monótona creciente, para obtener una estimación de P de manera que sea independiente de p , debemos tomar el menor valor posible (para los distintos p) de T . Tal mínimo se alcanza para $p = 1/2$ y es igual a 4. Así, aproximadamente

$$P \geq F(4) = 0,99993. \quad (19)$$

En la desigualdad (19) no se ha tenido en cuenta el error originado por el carácter aproximado de la fórmula (17). Estimando este error, podemos probar que en cualquier caso $P > 0,9999$.

En relación con este ejemplo de aplicación de la fórmula (17), diremos que las estimaciones del término restante de la fórmula (17) dadas en trabajos teóricos sobre la teoría de probabilidades fueron insatisfactorias durante mucho tiempo. Así, las aplicaciones de (17) y fórmulas similares a cálculos basados en valores pequeños de n , o con probabilidades p muy próximas a cero o uno (tales probabilidades son, con frecuencia, de particular importancia) estaban a menudo basadas sólo en verificaciones experimentales de resultados de esta

clase para un restringido número de ejemplos, y no en cualquier estimación válida del posible error. Asimismo, se probó mediante una investigación más detallada que, en muchos casos prácticos importantes, las fórmulas asintóticas introducidas anteriormente requieren no sólo una estimación del término restante, sino también ciertos refinamientos ulteriores (sin los cuales este término sería demasiado grande). Los resultados más completos en ambas direcciones se deben a S. N. Bernstein.

Las relaciones (11), (17) y (18) se pueden escribir en la forma

$$P \left\{ \left| \frac{\mu}{n} - p \right| < t \sqrt{\frac{p(1-p)}{n}} \right\} \sim F(t). \quad (20)$$

Para t suficientemente grande, el segundo miembro de la fórmula (20), que no contiene a n , es arbitrariamente próximo a uno, es decir, al valor de la probabilidad que da certeza absoluta. Vemos, de esta forma, que, *por regla general, la desviación de la frecuencia μ/n respecto de la probabilidad, es de orden $1/\sqrt{n}$* . Tal proporcionalidad entre la exactitud de una ley de probabilidad y la raíz cuadrada del número de observaciones es típica de otras muchas cuestiones. A veces se dice incluso en simplificaciones populares que «la ley de la raíz cuadrada de n » es la ley básica de la teoría de probabilidades. Se alcanzó una precisión completa acerca de esta idea gracias a la introducción y uso sistemático, por el gran matemático ruso P. L. Chebichev, de los conceptos de «esperanza matemática» y «varianza» para sumas y medias aritméticas de «variables aleatorias».

Variable aleatoria es el nombre dado a una cantidad que bajo condiciones dadas S puede tomar distintos valores con probabilidades específicas. A nosotros nos basta considerar variables aleatorias que tomen solamente un número finito de valores distintos. Para dar la *distribución de probabilidad*, como se llama, de una variable aleatoria de este tipo ξ , basta indicar sus posibles valores $x_1, x_2, \dots, x_n$ y las probabilidades

$$P_r = P \{ \xi = x_r \}.$$

La suma de estas probabilidades para todos los valores posibles de la variable ξ es siempre igual a uno:

$$\sum_{r=1}^s P_r = 1.$$

El número de resultados positivos en n experimentos puede servir como ejemplo de una variable aleatoria.

La *esperanza matemática* de la variable ξ es la expresión

$$M(\xi) = \sum_{r=1}^s P_r x_r,$$

y la *varianza* de ξ es la esperanza matemática del cuadrado de la desviación $\xi - M(\xi)$, es decir, la expresión

$$D(\xi) = \sum_{r=1}^s P_r [x_r - M(\xi)]^2.$$

La raíz cuadrada de la varianza

$$\sigma_\xi = \sqrt{D(\xi)}$$

se denomina *desviación típica* —de la variable a partir de su esperanza matemática $M(\xi)$.

La famosa desigualdad de Chebichev

$$P\{|\xi - M(\xi)| \leq t\sigma_\xi\} \geq 1 - \frac{1}{t^2}. \quad (21)$$

es fundamental en las aplicaciones más simples de la varianza y la desviación típica. Revela que las desviaciones de ξ respecto de $M(\xi)$ significativamente mayores que σ_ξ son escasas.

En cuanto a la suma de variables aleatorias

$$\xi = \xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)},$$

sus esperanzas matemáticas verifican siempre la ecuación

$$M(\xi) = M(\xi^{(1)}) + M(\xi^{(2)}) + \dots + M(\xi^{(n)}). \quad (22)$$

Pero la ecuación análoga para la varianza

$$D(\xi) = D(\xi^{(1)}) + D(\xi^{(2)}) + \dots + D(\xi^{(n)}) \quad (23)$$

sólo es cierta bajo determinadas condiciones. Para la validez de la ecuación (23) es suficiente, por ejemplo, que no haya «correlación»

entre las variables $\xi^{(i)}$ y $\xi^{(j)}$ con índices distintos, es decir, que para $i \neq j$ se verifique la ecuación ⁸

$$\mathbf{M}\{\xi^{(i)} - \mathbf{M}(\xi^{(i)})\} [\xi^{(j)} - \mathbf{M}(\xi^{(j)})] = 0. \quad (24)$$

En particular, la ecuación (24) es válida si las variables $\xi^{(i)}$ y $\xi^{(j)}$ son independientes entre sí ⁹. Por consiguiente, para términos mutuamente independientes la ecuación (23) es siempre válida. Para la media aritmética

$$\zeta = \frac{1}{n} (\xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)})$$

se sigue de (23) que

$$\mathbf{D}(\zeta) = \frac{1}{n^2} [\mathbf{D}(\xi^{(1)}) + \mathbf{D}(\xi^{(2)}) + \dots + \mathbf{D}(\xi^{(n)})]. \quad (25)$$

Supongamos ahora que para cada uno de estos términos la varianza no excede de una cierta constante

$$\mathbf{D}(\xi^{(i)}) \leq C^2.$$

Entonces, en virtud de (25)

$$\mathbf{D}(\zeta) \leq \frac{C^2}{n},$$

⁸ El coeficiente de correlación entre las variables $\xi^{(i)}$ y $\xi^{(j)}$ es la expresión

$$R = \frac{\mathbf{M}\{[\xi^{(i)} - \mathbf{M}(\xi^{(i)})][\xi^{(j)} - \mathbf{M}(\xi^{(j)})]\}}{\sigma_{\xi^{(i)}} \sigma_{\xi^{(j)}}}.$$

Si $\sigma_{\xi^{(i)}} > 0$ y $\sigma_{\xi^{(j)}} > 0$, entonces la condición (24) es equivalente a decir que $R = 0$.

El coeficiente de correlación R caracteriza el grado de dependencia entre variables aleatorias. $|R| \leq 1$ siempre, y $R = \pm 1$ solamente para una relación lineal

$$\eta = a\xi + b \quad (a \neq 0).$$

Para variables independientes $R = 0$.

⁹ La independencia de dos variables aleatorias ξ y η , que pueden tomar, respectivamente, los valores $x_1, x_2, \dots, x_m$ e $y_1, y_2, \dots, y_n$, se define de forma que para cada i y j los sucesos $A_i = \{\xi = x_i\}$ y $B_j = \{\eta = y_j\}$ son independientes en el sentido de la definición dada en §2.

y teniendo en cuenta la desigualdad de Chebichev para cualquier t

$$P\left\{|\zeta - M(\zeta)| \leq \frac{tC}{\sqrt{n}}\right\} \geq 1 - \frac{1}{t^2}. \quad (26)$$

La desigualdad (26) expresa la llamada ley de los grandes números, en la forma establecida por Chebichev: si las variables $\xi^{(i)}$ son mutuamente independientes y tienen varianza acotada, entonces al aumentar n la media aritmética ζ se desvía cada vez menos de la esperanza matemática $M(\zeta)$.

Con más precisión, la *sucesión de variables*

$$\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}, \dots$$

se dice que obedece la ley de los grandes números si para la media aritmética ζ y para cualquier constante $\varepsilon > 0$

$$P\{|\zeta - M(\zeta)| \leq \varepsilon\} \rightarrow 1 \quad (27)$$

cuando $n \rightarrow \infty$.

Para pasar la desigualdad (26) a la relación límite (27) basta poner

$$t = \varepsilon \frac{\sqrt{n}}{C}.$$

A. A. Markov, S. N. Bernstein, A. Ja. Hinchin y otros consagraron un gran número de investigaciones a ampliar tanto como fuera posible las condiciones bajo las cuales la relación límite es válida, es decir, las condiciones para la validez de la ley de los grandes números. Estas investigaciones tienen un significado teórico fundamental, pero aún más importante es un estudio exacto de la distribución de probabilidad para la variable $\zeta - M(\zeta)$.

Uno de los mayores servicios prestados por la escuela rusa clásica de matemáticos a la teoría de probabilidades es el establecimiento de que, bajo condiciones muy amplias, la ecuación

$$P\{t_1\sigma_\zeta < \zeta - M(\zeta) < t_2\sigma_\zeta\} \sim \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2} dt \quad (28)$$

es asintóticamente válida (es decir, con mayor y mayor exactitud a medida que n crece indefinidamente).

Chebichev dio una demostración casi completa de esta fórmula para el caso de términos independientes y acotados. Markov llenó una laguna en el argumento de Chebichev y amplió las condiciones de aplicabilidad de (28). Liapunov dio aún condiciones más generales. La aplicabilidad de (28) a la suma de términos mutuamente independientes fue estudiada de una manera particularmente exhaustiva por S. N. Bernstein.

La fórmula (28) abarca un número tan grande de casos particulares que se ha denominado mucho tiempo teorema central del límite en teoría de probabilidades. Aun cuando se ha probado recientemente que se debe incluir en una serie de leyes más generales, su valor difícilmente se podrá exagerar incluso en el momento presente.

Si las variables son independientes y sus varianzas son todas iguales a

$$D(\xi^{(i)}) = \sigma^2,$$

entonces conviene utilizar (25) y poner la fórmula (28) en la forma

$$P \left\{ \frac{t_1 \sigma}{\sqrt{n}} < \zeta - M(\zeta) < \frac{t_2 \sigma}{\sqrt{n}} \right\} \sim \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2} dt. \quad (29)$$

Probemos que la relación (29) contiene la solución del problema, considerado anteriormente, de calcular la desviación de la frecuencia μ/n respecto de la probabilidad p . Para ello introducimos las variables aleatorias $\xi^{(i)}$, definidas como sigue:

$$\xi^{(i)} = \begin{cases} 0, & \text{si la } i\text{-ésima prueba da un resultado negativo,} \\ 1, & \text{si la } i\text{-ésima prueba da un resultado positivo.} \end{cases}$$

Es fácil deducir que entonces

$$\mu = \xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)}, \quad \frac{\mu}{n} = \zeta, \\ M(\xi^{(i)}) = p, \quad D(\xi^{(i)}) = p(1-p), \quad M(\zeta) = p,$$

y la fórmula (29) da

$$P \left\{ t_1 \sqrt{\frac{p(1-p)}{n}} < \frac{\mu}{n} - p < t_2 \sqrt{\frac{p(1-p)}{n}} \right\} \sim \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2} dt,$$

que para $t_1 = -t$, $t_2 = t$ lleva otra vez a la fórmula (20).

§4. *Observaciones adicionales sobre los conceptos básicos de la teoría de probabilidades*

Al hablar de sucesos aleatorios, que tienen la propiedad de que sus frecuencias tienden a estabilizarse —es decir, en una larga sucesión de experimentos repetidos bajo condiciones prefijadas, sus frecuencias se agrupan en torno a un *nivel standard*, llamado su probabilidad $P(A/S)$ — hemos pecado, en la sección 1, de cierta vaguedad en nuestras formulaciones, y ello en dos aspectos. En primer lugar, no indicamos cuántos experimentos debemos hacer para establecer, fuera de toda duda, la existencia de la supuesta estabilidad; en otras palabras, no dijimos qué desviaciones de las frecuencias μ_r/n_r (una respecto a otra o respecto a su nivel standard p) eran aceptables para sucesiones de pruebas $n_1, n_2, \dots, n_s$ de longitud dada. Esta inexactitud en la primera etapa de la formulación de conceptos de una ciencia nueva es inevitable. No es mayor que la conocida vaguedad que rodea a los conceptos geométricos más simples de punto y recta y su sentido *físico*. Este aspecto de la cuestión se aclaró en la sección 3.

Más fundamental, sin embargo, es la segunda falta de claridad de que adolecen nuestras formulaciones: se refiere a la manera de formar las sucesiones de ensayos en los que se pretende examinar la estabilidad de la frecuencia con que aparece el suceso A .

Como ya dijimos anteriormente, en aquellos casos en que sea imposible una predicción exacta del curso de los acontecimientos, es preciso recurrir a métodos estadísticos y probabilísticos de investigación. Pero si queremos formar de algún modo artificial una sucesión de sucesos que, dentro de lo posible, sea puramente aleatoria, entonces debemos tener especial cuidado de que no haya ningún método para determinar de antemano en qué casos es probable que ocurra A con más frecuencia de lo normal.

Tales precauciones se toman, por ejemplo, en la organización de loterías nacionales. Si en una determinada lotería de N boletos se van a premiar M , entonces la probabilidad de que un boleto dado salga premiado es igual a $p = M/N$. Esto significa que cualquiera que sea la forma en que seleccionemos, antes del sorteo, un conjunto suficientemente grande de n boletos, podamos estar prácticamente seguros de que la razón μ/n , entre el número μ de boletos premiados en el conjunto elegido y el número total n de boletos de dicho conjunto, será próxima a p . Por ejemplo, la gente que prefiera boletos

marcados con un número par no tendrá ninguna ventaja sistemática sobre quienes los prefieran con un número impar, y análogamente no habrá ninguna ventaja al proceder según el criterio, por ejemplo, de que siempre es mejor comprar boletos con números que tengan exactamente tres factores primos, o boletos cuyos números sean próximos a los premiados en la lotería anterior, etc.

Igualmente, cuando disparamos un cañón bien construido, manejado por un equipo bien adiestrado y con proyectiles que hayan sido sometidos a un control de calidad, la desviación a partir de la posición media de los impactos de los proyectiles será menor que la desviación probable B previamente determinada en la *mitad* de los casos aproximadamente. Esta proporción permanece igual en una serie de pruebas sucesivas, y también en caso de que contemos separadamente el número de desviaciones que son menores que B para disparos pares (numerándolos en el orden en que se efectúan) o para impares. Pero es perfectamente posible que si hacemos una selección de proyectiles particularmente homogéneos (con respecto al peso, etc.), la dispersión disminuya considerablemente; tendríamos entonces una sucesión de disparos para los que la fracción de las desviaciones mayores que la *standard* sería considerablemente menor de la mitad.

Por tanto, decir que un suceso A es «aleatorio» o «estocástico» y asignarle una probabilidad determinada

$$p = P(A/S)$$

solamente es posible si hemos determinado de antemano la clase de caminos lícitos para el montaje de la serie de experimentos. La naturaleza de esta clase se supondrá incluida en las condiciones S .

Para condiciones *dadas* S , las propiedades del suceso A de ser aleatorio y de tener la probabilidad $p = P(A/S)$ expresan el carácter objetivo de la relación entre la condición S y el suceso A . En otras palabras, no existe ningún suceso que sea absolutamente aleatorio; un suceso es aleatorio o predeterminado según el contexto en que se lo considere, pero bajo condiciones específicas un suceso puede ser aleatorio en un sentido totalmente no-subjetivo, es decir, independientemente de la información de cualquier observador. Si imaginamos un observador que sea capaz de dominar todas las propiedades específicas y circunstancias particulares de la trayectoria de proyectiles, y pueda de este modo predecir, para cada uno de ellos, la desviación de la trayectoria media, su presencia no evitará que los proyectiles se dispersen siguiendo las leyes de la teoría de probabili-

dades, siempre que, naturalmente, el disparo se haga de la manera usual, y no de acuerdo con las instrucciones de nuestro observador imaginario.

En este sentido, observamos que la formación de una serie del tipo discutido antes, en la que las frecuencias tienden a hacerse constantes en el sentido de agruparse en torno a un valor normal, llamado probabilidad, se realiza en el mundo real de una forma completamente independiente de nuestra intervención. Por ejemplo, es precisamente en virtud del carácter aleatorio del movimiento de las moléculas de un gas por lo que el número de las que (incluso en un intervalo de tiempo muy pequeño) golpean una pequeña sección arbitrariamente prefijada de la pared del recipiente (o de la superficie de cuerpos situados en el gas) es proporcional, con gran aproximación, al área de esta sección de pared y al tiempo. Las desviaciones respecto de esta proporcionalidad, cuando el número de impactos es pequeño, también siguen las leyes de la teoría de probabilidades y producen fenómenos del tipo del movimiento browniano, del que hablaremos más adelante.

Volvamos ahora al sentido objetivo del concepto de independencia. Recordemos que la probabilidad condicional de un suceso A bajo la condición B se define mediante la fórmula

$$P(A/B) = \frac{P(AB)}{P(B)}. \quad (30)$$

Recordemos también que los sucesos A y B se dicen independientes si, como en (4),

$$P(AB) = P(A) P(B).$$

De la independencia de los sucesos A y B y del hecho de que $P(B) > 0$ se deduce que

$$P(A/B) = P(A).$$

Todos los teoremas de la teoría matemática de probabilidades que tratan de sucesos independientes se aplican a cualquier suceso que verifique la condición (4), o su generalización al caso de independencia mutua de varios sucesos. Sin embargo, estos teoremas tendrán poco interés si la definición no guarda relación alguna con las

propiedades de sucesos objetivos que son independientes en el sentido causal.

Se sabe, por ejemplo, que la probabilidad de dar a luz a un varón es, con suficiente estabilidad, $P(A) = 22/43$. Si B denota la condición de que el parto ocurra en la conjunción de Júpiter con Marte, entonces, bajo la hipótesis de que los planetas no influyen en el destino de los individuos, la probabilidad condicional $P(A/B)$ tiene el mismo valor: $P(A/B) = 22/43$; es decir, el cálculo efectivo de la frecuencia de nacimientos de varones bajo tales condiciones astrológicas especiales dará precisamente la misma frecuencia $22/43$. Aunque es probable que tal cálculo no se haya hecho nunca en una escala suficientemente grande, no hay razón para poner en duda el resultado.

Damos este ejemplo, de un tema algo anticuado, para mostrar que el desarrollo del conocimiento humano no solamente consiste en establecer relaciones válidas entre fenómenos, sino también en refutar relaciones imaginadas y establecer en casos pertinentes la tesis de la independencia de dos conjuntos cualesquiera de sucesos. Este desenmascaramiento de los absurdos esfuerzos de los astrólogos por relacionar dos clases de sucesos que de hecho no lo están, es uno de los ejemplos clásicos.

Naturalmente, al tratar del concepto de independencia no debemos proceder de una manera demasiado absoluta. Por ejemplo, debido a la ley de gravitación universal, es un hecho indudable que los movimientos de los satélites de Júpiter tienen un cierto efecto, digamos, sobre la trayectoria de un proyectil de artillería. Pero también es evidente que esta influencia se puede ignorar en la práctica. Desde el punto de vista filosófico, sería más propio hablar no de la independencia, sino de la insignificancia de la dependencia de ciertos sucesos. Como quiera que sea, la independencia de sucesos en el sentido relativo y concreto de este término no contradice en modo alguno el principio de la interconexión universal de todos los fenómenos; sirve solamente de complemento necesario a este principio.

El cálculo de probabilidades mediante fórmulas deducidas bajo la hipótesis de la independencia de ciertos sucesos tiene interés práctico incluso en aquellos casos en que los sucesos, aunque originariamente independientes, se hagan dependientes como resultado de los sucesos mismos. Por ejemplo, las probabilidades de colisión de partículas de radiación cósmica con partículas del medio atravesado por la radiación se pueden calcular suponiendo que hasta el momento de la aparición de una partícula de radiación cósmica, el movimiento

de las partículas del medio prosigue independientemente del movimiento de dicha partícula cósmica. Asimismo, la probabilidad de que una bala enemiga golpee la paleta de una hélice puede calcularse suponiendo que la posición de dicha paleta con respecto al eje de rotación no depende de la trayectoria de la bala, hipótesis que naturalmente será falsa para las balas del propio aviador, ya que éstas se disparan entre las paletas de la hélice. El número de ejemplos se podría extender sin límite.

Incluso cabe decir que allí donde aparecen leyes probabilísticas de un modo bien definido, nos encontramos ante la influencia de un gran número de factores que, si bien no son totalmente independientes unos de otros, están relacionados sólo en un sentido débil.

Esto no significa en absoluto que debamos introducir alegremente hipótesis de independencia. Por el contrario, nos obliga, en primer lugar, a proceder con suma cautela en la elección de criterios para comprobar las hipótesis de independencia, y en segundo lugar, a ser muy prudentes al investigar casos límite en los que se debe suponer la dependencia, aunque ésta es de tal clase que introduce complicaciones en las leyes de probabilidad relevantes. Ya indicamos anteriormente que la escuela rusa clásica de la teoría de probabilidades ha llevado a cabo amplias investigaciones en esta dirección.

Para terminar nuestra discusión sobre el concepto de independencia, observemos que, al igual que con la definición de independencia de dos sucesos dada en la fórmula (4), la definición formal de la independencia de varias variables aleatorias es mucho más amplia que el concepto de independencia en el mundo práctico, es decir, la ausencia de relación causal.

Supongamos, por ejemplo, que el punto ξ cae en el intervalo $[0,1]$ de tal manera que para

$$0 \leq a \leq b \leq 1$$

la probabilidad de que pertenezca al segmento $[a, b]$ es igual a la longitud del mismo $b - a$. Es fácil probar que en el desarrollo

$$\xi = \frac{\alpha_1}{10} + \frac{\alpha_2}{100} + \frac{\alpha_3}{1000} + \dots$$

de la abscisa del punto ξ en una fracción decimal, los dígitos α_k son mutuamente independientes, aunque están relacionados por la forma

en que se obtienen¹⁰. (De este hecho se deducen muchos resultados teóricos, algunos de los cuales tienen interés práctico.)

Tal flexibilidad en la definición formal de independencia no debe contemplarse como un defecto. Por el contrario, solamente extiende el dominio de aplicabilidad de teoremas establecidos para una u otra hipótesis de independencia. Estos teoremas son tan aplicables a casos en que la independencia se postule sobre la base de consideraciones prácticas como a casos en que la independencia se pruebe mediante cálculos efectuados a partir de supuestos previos relativos a las distribuciones de probabilidad de los sucesos y a las variables aleatorias en consideración.

En general, la investigación de la estructura formal del aparato matemático de la teoría de probabilidades ha llevado a interesantes resultados. Este aparato ocupa un lugar muy definido y preciso en la clasificación —hoy día cada vez más clara en sus líneas generales— de los objetos básicos de estudio de la matemática actual.

Yá hemos hablado de los conceptos de intersección AB y unión $A \cup B$ de los sucesos A y B . Recordemos que dos sucesos se dicen mutuamente excluyentes si su intersección es vacía, o lo que es igual, si $AB = N$, donde N es el símbolo utilizado para un suceso imposible.

El axioma básico de la teoría elemental de probabilidades consiste en exigir (véase §2) que bajo la condición $AB = N$ se verifique la ecuación

$$P(A \cup B) = P(A) + P(B).$$

Los conceptos fundamentales de la teoría de probabilidades, esto es, los sucesos y sus probabilidades, son completamente análogos en sus propiedades a las figuras planas y sus áreas. Basta entender por AB la intersección (parte común) de dos figuras; por $A \cup B$, su unión; por N , la figura «vacía» convencional, y por $P(A)$, el área de la figura A ; la analogía es entonces completa.

Las mismas observaciones sirven para los volúmenes de figuras tridimensionales.

La teoría más general de entes de este tipo, que contiene como casos especiales a la de volúmenes y áreas, se suele denominar *teoría de la medida*, discutida en el capítulo 14 en relación con la teoría de funciones de una variable real.

¹⁰ Esto también es válido, para cualquier n , para los dígitos a_k del desarrollo del número ξ en la fracción

$$\xi = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots$$

Queda solamente observar que la teoría de probabilidades —en comparación con la teoría general de la medida o en particular con la de áreas y volúmenes— exhibe cierto rasgo especial: una probabilidad nunca es mayor que uno. Esta probabilidad maximal vale para un suceso necesario U .

$$P(U) = 1.$$

La analogía no es de ningún modo superficial, pues se comprueba que toda la teoría matemática de probabilidades se puede construir, desde el punto de vista formal, como una teoría de la medida, haciendo la hipótesis especial de que la medida de «todo el espacio» U es igual a uno ¹¹.

Tal enfoque del tema deja completamente clara la construcción formal de la teoría matemática de probabilidades, y ha permitido un progreso sustancial no solamente en dicha teoría, sino también en otras estrechamente relacionadas con ella en cuanto a su estructura formal. En la teoría de probabilidades, el éxito se ha obtenido a través de métodos refinados desarrollados en la teoría métrica de funciones de una variable real; y al mismo tiempo, los métodos probabilísticos han demostrado ser aplicables a cuestiones de otros dominios vecinos de las matemáticas, no «por analogía», sino por un traslado estricto y formal de ellos al nuevo dominio. Dondequiera que probemos que los axiomas de la teoría de probabilidades se verifican, los resultados de éstos son aplicables, aun cuando el dominio dado no tenga nada que ver con el azar en el mundo real.

La existencia de una teoría de probabilidades axiomatizada nos preserva de la tentación de «definir» la probabilidad por métodos que aspiran a construir una teoría matemática estricta, puramente formal, sobre la base de rasgos de la probabilidad sugeridos directamente por las ciencias naturales. Tales definiciones corresponderían en líneas generales a la «definición» en geometría de un punto como el resultado de partir un cuerpo por la mitad infinitas veces.

Con definiciones de esta clase, la probabilidad se toma como el límite de la frecuencia cuando el número de experimentos crece indefinidamente. La misma hipótesis de que los experimentos son probabilísticos, esto es, que las frecuencias tienden a agruparse en torno

¹¹ No obstante, debido a la naturaleza de sus problemas, la teoría de probabilidades sigue siendo una disciplina matemática independiente; sus resultados básicos (presentados con detalle en §3) parecen artificiales e innecesarios desde el punto de vista de la teoría de la medida pura.

a un valor constante, sólo será válida (y lo mismo ocurre para la «aleatoriedad» de un suceso particular) si ciertas condiciones se mantienen fijas por tiempo ilimitado y con exactitud absoluta. Por tanto, el paso al límite

$$\frac{\mu}{n} \rightarrow p$$

no puede tener ningún sentido objetivo. La formulación del principio de estabilidad de las frecuencias en tal proceso límite exige que definamos los métodos lícitos de construir una sucesión infinita de experimentos, y esto solamente puede hacerse con una ficción matemática. Toda esta aglomeración de conceptos podría merecer una consideración seria si el resultado final fuese una teoría de tal naturaleza que no existiera otro medio de colocarla sobre una base rigurosa. Pero, como ya dijimos antes, la teoría matemática de probabilidades se puede basar en la de la medida (en su forma actual) añadiendo simplemente la condición

$$P(U) = 1.$$

En general, para cualquier análisis práctico del concepto de probabilidad no es necesario referirse a su definición formal. Es obvio que en lo que concierne al lado puramente formal de la probabilidad, solamente podemos decir lo siguiente: la probabilidad $P(A/S)$ es un número en torno al cual, bajo condiciones S que determinan la manera lícita de montar los experimentos, tienden a agruparse las frecuencias; que esta tendencia se cumplirá con una exactitud cada vez mayor a medida que crece el número de experimentos (siempre realizados de tal forma que se conserven las condiciones originales); y finalmente, que la tendencia alcanzará un grado satisfactorio de fiabilidad y exactitud tras un número razonable de experimentos.

De hecho, el problema de importancia, en la práctica, no es el de dar formalmente una definición precisa de aleatoriedad, sino aclarar en lo posible las condiciones bajo las que se da la aleatoriedad del tipo citado. Hay que tener muy presente que, en realidad, las hipótesis relativas al carácter probabilístico de un fenómeno raramente están basadas en una comprobación estadística directa. Solamente es en la primera etapa de la introducción de métodos estadísticos en un nuevo dominio de la ciencia cuando el trabajo consiste en una observación puramente empírica de la constancia de las frecuencias. En virtud de la sección 3, vemos que la comprobación estadística de la

constancia de las frecuencias con una aproximación ϵ requiere una serie de experimentos, consistente cada uno en $n = 1/\epsilon^2$ pruebas. Por ejemplo, para establecer que en un problema concreto la probabilidad está determinada con una aproximación de 0,0001, es necesario efectuar una serie de experimentos que contengan aproximadamente 100.000.000 de pruebas cada uno.

Mucho más frecuente es que la hipótesis de aleatoriedad probabilística se introduzca por consideraciones de simetría o de series sucesivas de sucesos, con la subsiguiente comprobación de las hipótesis de una forma indirecta. Por ejemplo, como el número de moléculas en un volumen finito de un gas es del orden de 10^{20} o más, el número $\sqrt{n}$, correspondiente a las deducciones probabilísticas hechas en la teoría cinética de gases, es muy grande, de suerte que muchas de estas deducciones se verifican con gran exactitud. Así, las presiones sobre los lados opuestos de una lámina suspendida en aire inmóvil, aunque la lámina sea de dimensiones microscópicas, son exactamente iguales, y cualquier exceso de presión, a un lado u otro, del orden de una milésima por ciento se puede detectar en un experimento dispuesto convenientemente.

§5. *Procesos determinados y aleatorios*

El principio de relación causal entre fenómenos encuentra su expresión matemática más simple en el estudio de procesos físicos por medio de ecuaciones diferenciales, como se ha demostrado en una serie de ejemplos en la sección 1 del capítulo 5.

Supongamos que el estado del sistema en estudio está determinado en el instante t por n parámetros

$$x_1, x_2, \dots, x_n.$$

La variación de estos parámetros se expresa por sus derivadas con respecto al tiempo

$$\dot{x}_k = \frac{dx_k}{dt}.$$

Si se supone que estas variaciones son funciones de los valores de

los parámetros, entonces obtenemos un sistema de ecuaciones diferenciales

$$\dot{x}_1 = f_1(x_1, x_2, \dots, x_n),$$

$$\dot{x}_2 = f_2(x_1, x_2, \dots, x_n),$$

$$\dots\dots\dots$$

$$\dot{x}_n = f_n(x_1, x_2, \dots, x_n).$$

La mayor parte de las leyes de la naturaleza descubiertas en el momento del nacimiento de la física matemática, empezando por la ley de Galileo para la caída de los graves, están expresadas de esta manera precisamente. Galileo no pudo expresar su descubrimiento en esta forma, ya que en su época no habían sido desarrollados aún los conceptos matemáticos correspondientes; Newton fue el primero en hacerlo.

En mecánica y en otros campos de la física se acostumbra a expresar estas leyes mediante ecuaciones diferenciales de segundo orden. Pero en ello no interviene ningún principio nuevo; pues si denotamos los $\dot{x}_k$ por los nuevos símbolos

$$v_k = \dot{x}_k,$$

obtenemos para la segunda derivada de las cantidades x_k las expresiones

$$\frac{d^2 x_k}{dt^2} = \dot{v}_k,$$

y las ecuaciones de segundo orden en las n cantidades $x_1, x_2, \dots, x_n$ se convierten en otras de primer orden en las $2n$ cantidades $x_1, x_2, \dots, x_n, v_1, v_2, \dots, v_n$.

Consideremos, como un ejemplo, la caída de un cuerpo pesado en la atmósfera de la tierra. Si consideramos solamente pequeñas distancias por encima de la superficie, podemos suponer que la resistencia del medio depende sólo de la velocidad y no de la altura. El estado del medio en cuestión está caracterizado por dos parámetros: la distancia z del cuerpo a la superficie de la tierra y su velocidad v . La variación de estas dos cantidades con el tiempo viene dada por las dos ecuaciones diferenciales

$$\begin{aligned} \dot{z} &= -v, \\ \dot{v} &= g - f(v), \end{aligned} \tag{31}$$

donde g es la aceleración de la gravedad y $f(v)$ es cierta «ley de resistencia» para el cuerpo dado.

Si la velocidad no es grande y el cuerpo es suficientemente masivo —digamos una piedra de medianas dimensiones cayendo desde una altura de varios metros—, la resistencia del aire se puede despreciar y las ecuaciones (31) se transforman en las siguientes:

$$\begin{aligned}\dot{z} &= -v, \\ \dot{v} &= g.\end{aligned}\tag{32}$$

Si se supone que en el instante inicial t_0 las cantidades z y v toman los valores z_0 y v_0 , entonces es fácil resolver las ecuaciones (32), obteniendo la fórmula

$$z = z_0 - v(t - t_0) - g \left(\frac{t - t_0}{2} \right)^2,$$

que describe el proceso total de caída. Por ejemplo, si $t_0 = 0$, $v_0 = 0$ obtenemos

$$z = z_0 - \frac{gt^2}{2},$$

fórmula hallada por Galileo.

En el caso general, la integración de las ecuaciones (31) es más difícil, si bien el resultado fundamental, con restricciones muy generales sobre la función $f(v)$, sigue siendo el mismo: conocidos los valores z_0 y v_0 en el instante inicial t_0 , los valores de z y v en todos los instantes ulteriores t están unívocamente determinados hasta el momento en que el cuerpo golpea la superficie de la tierra. Teóricamente, esta última restricción también se puede evitar, si suponemos que la caída se extiende a valores negativos de z . Para problemas presentados de esta manera se puede establecer lo siguiente: si la función $f(v)$ es monótona para v creciente y tiende a infinito cuando $v \rightarrow \infty$, entonces, si la caída continúa libre (cuando t aumenta indefinidamente), la velocidad v tiende a un valor límite constante c , que es la solución de la ecuación

$$g = f(c).$$

Desde un punto de vista intuitivo, este resultado del análisis matemático del problema es bastante comprensible: la velocidad de

caída aumenta hasta el momento en que la fuerza aceleratriz de la gravedad se ve contrarrestada por la resistencia del aire. En un descenso con paracaídas, la velocidad estacionaria v de cinco metros por segundo aproximadamente se alcanza más aprisa¹². En un salto desde gran altura en caída libre la resistencia del aire es menor, por lo que la velocidad estacionaria es mayor y solamente se alcanza después de que el paracaidista ha recorrido un largo camino.

Para la caída de cuerpos ligeros, como una pluma zarandeada en el aire o un trocito de pelusa, el período inicial de aceleración es muy corto, a menudo casi inapreciable. El valor estacionario de caída queda establecido con mucha rapidez, y para una aproximación standard podemos considerar que durante toda la caída $v = c$. En este caso tenemos solamente la ecuación diferencial

$$\dot{z} = -c,$$

que se integra muy fácilmente:

$$z = z_0 - c(t - t_0).$$

Así es como caerá un pedacito de pelusa en aire completamente inmóvil.

Esta concepción determinista está tratada de manera completamente general en la teoría actual de sistemas dinámicos, a la que está dedicada una serie de importantes trabajos de matemáticos soviéticos, N. N. Bogoljubov, V. V. Stepanov y muchos otros. Esta teoría general incluye también como caso especial la formulación matemática de fenómenos físicos en los que el estado de un sistema no está definido por un número finito de parámetros como en el caso anterior, sino por una o más funciones; así ocurre, por ejemplo, en la mecánica de medios continuos. En tales casos, las leyes elementales de cambio de estado en intervalos de tiempo «infinitamente pequeños» están dadas, no por ecuaciones diferenciales ordinarias, sino por ecuaciones diferenciales en derivadas parciales o por algún otro medio. Pero los rasgos comunes a todas las formulaciones matemáticas deterministas de procesos reales son: primero, que el estado del sistema en estudio se considera completamente determinado por un ente matemático ω (un conjunto de n números reales, una o más

¹² Esta afirmación se debe tomar en el sentido de que en la práctica v alcanza pronto un valor realmente próximo a c .

funciones, etc.); y segundo, que los valores correspondientes a instantes $t > t_0$ están unívocamente determinados por el valor ω_0 en el instante inicial t_0

$$\omega = F(t_0, \omega_0, t).$$

Para fenómenos descritos por ecuaciones diferenciales, el proceso de hallar la función ϕ consiste, como hemos visto, en integrar estas ecuaciones con las condiciones iniciales $\omega = \omega_0$ para $t = t_0$.

Los partidarios del materialismo mecanicista suponían que tal formulación es una expresión exacta y directa del carácter determinista de los fenómenos reales, o del principio físico de causalidad. Según Laplace, el estado del mundo en un instante dado está definido por un número infinito de parámetros, sujetos a una infinidad de ecuaciones diferenciales. Si alguna «mente universal» pudiera anotar todas estas ecuaciones e integrarlas, podría predecir con absoluta precisión, según Laplace, la evolución total del mundo en el futuro infinito.

Pero de hecho este infinito matemático cuantitativo es extremadamente tosco en comparación con el carácter cualitativamente inagotable del mundo real. Ni la introducción de un número infinito de parámetros, ni la descripción del estado de medios continuos por funciones de un punto en el espacio es adecuada para representar la complejidad infinita de los sucesos reales.

Como ya indicamos en §3 del capítulo 5, el estudio de sucesos reales no siempre procede en el sentido de aumentar el número de parámetros introducidos en el problema; en general, dista mucho de ser conveniente complicar la ω que describe los distintos «estados del sistema» en nuestro esquema matemático. El arte de la investigación consiste más bien en encontrar un espacio muy simple Ω (un conjunto de valores de ω o, en otras palabras, de distintos estados posibles del sistema)¹³ tal que reemplazando el proceso real por variaciones del punto ω de una forma determinada sobre este espacio, podamos abarcar todos los aspectos *esenciales* del proceso real.

Pero si de un proceso real abstraemos sus aspectos esenciales, dejamos un residuo que debemos considerar aleatorio. Los factores aleatorios despreciados siempre ejercen cierta influencia sobre el curso de los procesos. Entre los fenómenos que admiten investigación matemática hay muy pocos que no muestren la influencia de

¹³ En el ejemplo dado antes de un cuerpo que cae, el espacio de las fases es el sistema de pares de números (z, v) , es decir, un plano. Para espacios de fases en general, véanse los capítulos 16 y 17.

factores aleatorios ignorados al comparar la teoría con la observación. Este es más o menos el estado de cosas en la teoría de movimientos planetarios sometidos a la fuerza de gravitación: la distancia entre los planetas es tan grande en comparación con su tamaño que su representación ideal como puntos materiales es casi completamente satisfactoria; el espacio en que se mueven está lleno de un material tan disperso que su resistencia a los movimientos es extremadamente pequeña; las masas de los planetas son tan grandes que la presión de la luz casi no juega ningún papel en sus movimientos. Estas circunstancias excepcionales explican el hecho de que la solución matemática para el movimiento de un sistema de n puntos materiales, cuyos «estados» están descritos por $6n$ parámetros¹⁴ que solamente tienen en cuenta la fuerza de la gravedad, coincide tan bien con la observación del movimiento de los planetas.

Un caso similar del movimiento planetario es la trayectoria de un proyectil de artillería sometido a la gravedad y a la resistencia del aire. Este es también uno de los campos clásicos en que los métodos matemáticos de investigación son relativamente fáciles y proporcionaron en seguida grandes éxitos. Pero aquí el papel de los factores aleatorios perturbadores es significativamente mayor y la dispersión de los proyectiles —su desviación de la trayectoria teórica— alcanza decenas de metros, o para grandes distancias incluso centenas de metros. Estas desviaciones están causadas en parte por desviaciones aleatorias en la dirección y velocidad iniciales, en parte por desviaciones aleatorias de la masa y el coeficiente de resistencia del proyectil, y en parte por ráfagas y otras irregularidades del viento y demás factores aleatorios que gobiernan las condiciones extraordinariamente complicadas y variables de la atmósfera de la tierra.

La dispersión de proyectiles se estudia con detalle por métodos de la teoría de probabilidades, y los resultados de este estudio son esenciales para la práctica de artillería.

¿Qué sentido tiene, propiamente hablando, el estudio de sucesos aleatorios? Podría pensarse que, cuando el «residuo» aleatorio para una formulación dada de un fenómeno es tan grande que no se puede despreciar, entonces el único modo posible de proceder sería describir el fenómeno con mayor aproximación introduciendo nuevos parámetros y hacer un estudio más detallado por el mismo método anterior.

Pero en muchos casos tal procedimiento es prácticamente irreali-

¹⁴ Las tres coordenadas y las tres componentes de la velocidad en cada punto.

zable. Por ejemplo, al estudiar la caída de un cuerpo material en la atmósfera, considerando el flujo irregular y borrascoso (o, como se suele llamar, turbulento) del viento, deberíamos introducir, en lugar de los dos parámetros z y v , un aparato matemático muy farragoso para describir esta estructura en su totalidad.

Pero, de hecho, este complicado procedimiento solamente es necesario en aquellos casos en que por alguna razón debemos determinar la influencia de estos factores residuales «aleatorios» con todo detalle y separadamente para cada factor individual. Afortunadamente, nuestras exigencias prácticas son generalmente muy diferentes; solamente necesitamos estimar el efecto total ejercido por los factores aleatorios para un largo intervalo de tiempo o para un gran número de repeticiones del proceso en estudio.

Consideremos, por ejemplo, el desplazamiento de la arena en el lecho de un río, o en una construcción hidroeléctrica. Normalmente este desplazamiento sucede de tal manera que la mayor parte de la arena permanece quieta, mientras que sólo de vez en vez una turbulencia especialmente fuerte próxima al fondo recoge granos individuales y los lleva a considerable distancia, donde quedan depositados de nuevo. El movimiento puramente teórico de cada grano se puede calcular individualmente por las leyes de la hidrodinámica, pero para esto haría falta determinar el estado inicial del fondo y del flujo en todos sus detalles y calcular el flujo paso a paso, observando aquellos instantes en que la presión sobre un grano particular de arena es suficiente para ponerlo en movimiento, y seguir la pista de este movimiento hasta que de pronto termine. Lo absurdo de plantear tal problema para estudios científicos reales es evidente. No obstante, las leyes medias, o como se suelen denominar, las leyes estadísticas del cambio de posición de la arena sobre el fondo del río son perfectamente susceptibles de investigación.

Ejemplos de este tipo, en que el efecto de un gran número de factores aleatorios conduce a leyes estadísticas completamente bien definidas, se podrían multiplicar fácilmente. Uno de los más conocidos y al mismo tiempo más fascinante, en vista de la amplitud de sus aplicaciones, es la teoría cinética de los gases, que muestra cómo la influencia conjunta de colisiones aleatorias de moléculas da lugar a leyes precisas que gobiernan la presión de un gas sobre la pared, la difusión de un gas en el seno de otro, etc.

§6. *Procesos aleatorios de tipo Markov*

A. A. Markov es el autor de un esquema probabilístico que viene a ser una generalización inmediata del esquema determinista del §5 descrito por la ecuación

$$\omega = F(t_0, \omega_0, t).$$

Es cierto que Markov consideró solamente el caso en que el espacio de las fases del sistema consiste en un número finito de estados $\Omega = (\omega_1, \omega_2, \dots, \omega_n)$ y que estudió el cambio de estado del sistema para cambios discretos del tiempo t solamente. Pero con este modelo estremadamente esquemático logró establecer una serie de leyes fundamentales.

En lugar de una función F que determine unívocamente el estado ω en el instante $t > t_0$ correspondiente al estado ω_0 en el instante t_0 , Markov introdujo las probabilidades

$$P(t_0, \omega_i; t, \omega_j)$$

de obtener el estado ω_j en el instante t con la condición de que en t_0 tengamos el estado ω_i . Estas probabilidades están conectadas para tres instantes cualesquiera.

$$t_0 < t_1 < t_2$$

por una relación, introducida por Markov, que se puede denominar la ecuación básica para un proceso de Markov

$$P(t_0, \omega_i; t_2, \omega_j) = \sum_{k=1}^n P(t_0, \omega_i; t_1, \omega_k) P(t_1, \omega_k; t_2, \omega_j). \quad (33)$$

Cuando el espacio de las fases es una variedad continua, el caso más típico es que exista una densidad de probabilidad $p(t_0, \omega_0; t, \omega)$ para pasar del estado ω_0 al ω en el intervalo de tiempo (t_0, t) . En este caso, la probabilidad de pasar del estado ω_0 a cualquiera de los ω pertenecientes a un dominio G del espacio de las fases Ω se puede escribir en la forma

$$P(t_0, \omega_0; t, G) = \int_G p(t_0, \omega_0; t, \omega) d\omega, \quad (34)$$

donde $d\omega$ es un elemento de volumen en el espacio de la fase¹⁵. Para la densidad de probabilidad $p(t_0, \omega_0; t, \omega)$, la ecuación básica (33) toma la forma

$$p(t_0, \omega_0; t_2, \omega_2) = \int_{\Omega} p(t_0, \omega_0; t_1, \omega) p(t_1, \omega; t_2, \omega_2) d\omega. \quad (35)$$

Normalmente la ecuación (35) es difícil de resolver, pero bajo restricciones conocidas podemos deducir de ella ciertas ecuaciones en derivadas parciales que son sencillas de estudiar. Algunas de estas ecuaciones fueron deducidas por los físicos Fokker y Planck mediante consideraciones físicas no rigurosas. Esta teoría de las llamadas ecuaciones diferenciales estocásticas fue construida en su forma completa por los autores soviéticos, S. N. Bernstein, A. N. Kolmogorov, I. G. Petrovski, A. Ja. Hinchin y otros. No daremos aquí estas ecuaciones.

El método de ecuaciones diferenciales estocásticas nos permite, por ejemplo, resolver sin dificultad el problema del movimiento en aire inmóvil de un cuerpo muy pequeño, para el que la velocidad media c de caída es mucho menor que la velocidad del «movimiento browniano» originado por el hecho de que sus colisiones con las moléculas de aire no están perfectamente equilibradas sobre sus distintos lados.

Sea c la velocidad media de caída, y D el llamado coeficiente de difusión. Si suponemos que una partícula no queda en la superficie de la tierra ($z = 0$) sino que se «refleja» (es decir, bajo la influencia de las fuerzas brownianas sale despedida otra vez hacia la atmósfera) y si suponemos también que en el instante t_0 la partícula está a una altura z_0 , entonces la densidad de probabilidad de estar a la altura z en el instante t viene expresada por la fórmula

$$\begin{aligned} p(t_0, z_0; t, z) = & \frac{1}{2\sqrt{\pi D(t-t_0)}} \times \\ & \times \left[e^{-\frac{(z-z_0)^2}{4D(t-t_0)}} + e^{-\frac{(z+z_0)^2}{4D(t-t_0)}} \right] e^{-\frac{(cz-z_0)}{2D} - \frac{c^2(t-t_0)}{4D}} + \\ & + \frac{c}{D\sqrt{\pi}} e^{-cz/D} \int_{\frac{z+z_0-c(t-t_0)}{2\sqrt{D(t-t_0)}}}^{\infty} e^{-z^2} dz. \end{aligned}$$

¹⁵ Hablando con propiedad, la ecuación (34) sirve para definir la densidad de probabilidad. La cantidad $p d\omega$ es igual (salvo un infinitésimo de orden superior) a la probabilidad de pasar, en el tiempo desde t_0 a t , del estado ω_0 al elemento de volumen $d\omega$.

En la figura 4 se ve cómo varían las curvas $p(t_0, z_0; t, z)$ para una sucesión de instantes t .

Vemos que la media de las alturas de la partícula aumenta, y su posición es cada vez más indeterminada, más «aleatoria». El aspecto más interesante de la situación es que para cualquier t_0 y z_0 y para $t \rightarrow \infty$

$$p(t_0, z_0; t, z) \rightarrow \frac{c}{D} e^{-cz/D}; \quad (36)$$

es decir, existe una distribución límite para la altura de la partícula, y la esperanza matemática de esta altura cuando crece t tiende a un límite positivo

$$z^* = \frac{c}{D} \int_0^{\infty} z e^{-cz/D} dz = \frac{D}{c}. \quad (37)$$

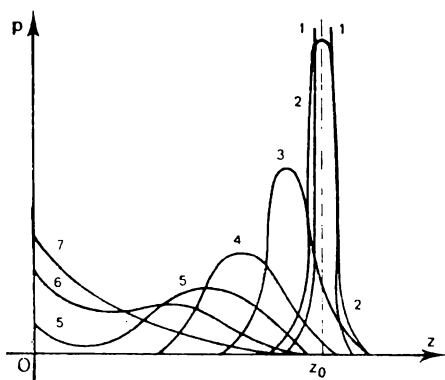


FIG. 4

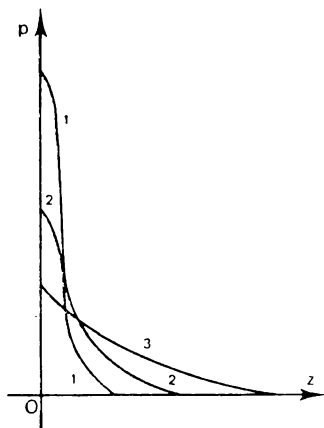


FIG. 5

Así, a pesar de que mientras nuestra partícula esté encima de la superficie de la tierra siempre tenderá a caer debido a la fuerza de la gravedad, sin embargo, como este proceso (el vagar por la atmósfera) continúa, la partícula se hallará por lo general a una altura positiva

determinada. Si tomamos el z_0 inicial menor que z^* , resultará que en un intervalo de tiempo suficientemente grande la posición media de la partícula estará más alta que su posición inicial, como se muestra en la figura 5, donde $z_0 = 0$.

Para partículas individuales los valores medios z^* en cuestión son solamente esperanzas matemáticas; pero, por la ley de los grandes números, se deduce que, para un gran número de partículas, dichas esperanzas se verán efectivamente realizadas: la densidad de distribución de la altura de tales partículas obedecerá las leyes indicadas, y, en particular, transcurrido un intervalo de tiempo suficiente, esta densidad se estabilizará de acuerdo con la fórmula (36).

Lo que se ha dicho hasta aquí es aplicable directamente sólo a los gases, al humo y sustancias similares que se hallan en el aire en pequeñas concentraciones, puesto que las cantidades c y D se suponían determinadas por un estado prefijado de la atmósfera. No obstante ciertas complicaciones, la teoría es aplicable a la difusión mutua de los gases que componen la atmósfera, y a la distribución en capas de sus densidades que se presenta como consecuencia de esta difusión.

El cociente c/D aumenta con el tamaño de las partículas, de modo que el carácter del movimiento cambia de difusión a caída libre de acuerdo con las leyes consideradas en el §5. La teoría nos permite descubrir todas las transiciones entre el movimiento puramente difusivo y tales leyes de caída.

El problema del movimiento de partículas suspendidas en una atmósfera turbulenta es más difícil, pero en principio se puede atacar con métodos probabilísticos similares.

§1. *Introducción*

En la vida práctica nos enfrentamos constantemente con el problema de aproximar ciertos números mediante otros. Por ejemplo, nuestras medidas de distintas magnitudes concretas, longitud, área, temperatura, etc., nos conducen a números que sólo son aproximaciones. En la práctica utilizamos solamente números racionales, números de la forma p/q , donde p y q ($q \neq 0$) son enteros. Pero además de los números racionales también existen los irracionales, y si bien no se utilizan para medir, no obstante nuestros argumentos teóricos conducen a menudo a ellos. Sabemos, por ejemplo, que la longitud de una circunferencia de radio $r = 1/2$ es igual al número irracional π , y la longitud de la hipotenusa de un triángulo rectángulo de lados unitarios es igual a $\sqrt{2}$. En cálculos efectivos con números irracionales, se suele utilizar, como primera aproximación, números racionales con el grado deseado de exactitud, normalmente por medio de una fracción decimal exacta.

La misma situación se da también para funciones. Las leyes cuantitativas de la naturaleza se expresan en matemáticas por medio de funciones, no con absoluta exactitud, sino aproximadamente, con distintos grados de precisión. Además, en un gran número de casos es necesario, incluso para funciones definidas por reglas completa-

mente matemáticas, aproximarlas por medio de otras funciones con una exactitud suficiente para poder calcularlas en la práctica.

Sin embargo, estas observaciones no se refieren al cálculo solamente. El problema de determinar una función por medio de otras tiene gran importancia teórica. Aclaremoslo en pocas palabras. El desarrollo del análisis matemático ha impulsado el descubrimiento y estudio de clases importantísimas de funciones que bajo condiciones conocidas han demostrado ser los medios naturales de aproximar otras más o menos arbitrarias. Estas clases resultaron ser, sobre todo, los polinomios algebraicos y trigonométricos, y también sus distintas generalizaciones. Se ha demostrado que a partir de las propiedades de la función que se desea aproximar es posible estimar, bajo ciertas condiciones, el carácter de su desviación respecto de una sucesión específica de funciones aproximadoras. Recíprocamente, si sabemos de qué modo se desvía de su aproximación mediante una sucesión de funciones, podemos establecer ciertas propiedades de la función. En esta dirección se ha construido una teoría de funciones basada sobre su representación aproximada mediante distintas clases de funciones aproximadoras. Existe una teoría similar en la teoría de números, y en ella se estudian las propiedades de los números irracionales apoyándose en sus aproximaciones mediante racionales.

En el capítulo 2 el lector ha encontrado ya un método importantísimo de aproximación: la fórmula de Taylor. Con su ayuda una función que satisfaga ciertas condiciones se aproxima con otra de la forma $P(x) = a_0 + a_1x + \cdots + a_nx^n$, llamada polinomio algebraico. Aquí los a_k son constantes, independientes de x .

Un polinomio algebraico tiene una estructura muy sencilla; para calcularlo para coeficientes conocidos a_k y valores dados de x , solamente necesitamos aplicar las tres operaciones aritméticas de adición, sustracción y multiplicación. La sencillez de este cálculo es extremadamente importante en la práctica y es una de las razones de que los polinomios algebraicos sean los medios más generalizados a la hora de aproximar funciones (otra razón importante se discutirá más adelante). Basta señalar que, especialmente en la actualidad, los cálculos técnicos se deben efectuar con máquinas computadoras a escala masiva. En el actual estado de perfección, las computadoras trabajan muy rápidamente, pero solamente pueden efectuar operaciones relativamente simples. Pueden realizar operaciones aritméticas con números muy altos, pero nunca, por ejemplo, procesos infinitos de paso al límite. Una máquina no puede calcular $\log x$ exactamente, pero lo que se puede hacer es aproximar esta función mediante un

polinomio $P(x)$ con el grado deseado de precisión, y luego calcular el polinomio con un computador.

Además de la fórmula de Taylor, hay otras de gran importancia práctica en la aproximación de funciones por polinomios algebraicos. Entre ellas están las distintas fórmulas de interpolación, muy utilizadas, en particular, en el cálculo aproximado de integrales, e incluso en la integración aproximada de ecuaciones diferenciales. Es bien conocido también el método de aproximación en el sentido del cuadrado medio, muy utilizado con otras funciones aparte de los polinomios algebraicos. Para ciertas cuestiones prácticas se atribuye gran importancia al método de aproximación uniforme óptima (o de Chebichev), ideado por el gran matemático ruso Chebichev y que surgió, como veremos, de la solución de un problema relacionado con la construcción de mecanismos.

Nuestro propósito aquí es dar al lector una idea de estos métodos y, hasta donde sea posible, establecer las condiciones bajo las cuales es preferible un método a otro. Ninguno de ellos es el mejor en términos absolutos. Cada uno de ellos es mejor que los demás bajo ciertas condiciones. Por ejemplo, si tenemos que resolver un problema físico, entonces alguno de los métodos de aproximar las funciones que aparecen en el problema vendrá particularmente indicado por el carácter mismo del problema o, como suele decirse, por consideraciones físicas. También veremos más adelante que, bajo condiciones bien conocidas, un método puede ser aplicable y otros no.

Cada uno de los métodos de cálculo surgió en su momento y tiene su teoría e historia características. Newton estaba ya familiarizado con una de las fórmulas de interpolación, que expresó de un modo muy conveniente para el cálculo práctico con los llamados cocientes de diferencias. El método de aproximación en el sentido del cuadrado medio tiene ciento cincuenta años por lo menos. Pero durante mucho tiempo estos métodos no dieron lugar a una teoría coherente. Eran solamente distintos métodos prácticos de aproximar funciones, y además, las limitaciones sobre su aplicabilidad no estaban claras.

La teoría actual de la aproximación de funciones emanó del trabajo de Chebichev, quien introdujo el importante concepto de aproximación óptima (y en particular de aproximación uniforme óptima), hizo uso sistemático de él en aplicaciones prácticas y desarrolló su base teórica. La aproximación óptima es el concepto fundamental de la moderna teoría de la aproximación. Después de Chebichev, sus ideas fueron desarrolladas más extensamente por sus

discípulos E. I. Zolotarev, A. N. Korkin, y los hermanos A. A. y V. A. Markov. En la época de Chebichev no sólo se introdujeron los conceptos fundamentales, sino que se encontraron métodos básicos para obtener la aproximación óptima de funciones concretas arbitrarias, métodos muy utilizados en la actualidad; también se hicieron investigaciones de las propiedades de las clases aproximadoras, concretamente de polinomios algebraicos y trigonométricos, desde el punto de vista de las necesidades surgidas de problemas prácticos.

El ulterior desarrollo de la teoría de aproximación de funciones se vio influido por un importante descubrimiento matemático, hecho a finales del siglo pasado por el matemático alemán Weierstrass, quien demostró con absoluto rigor la posibilidad de aproximar una función continua arbitraria por un polinomio algebraico con cualquier grado de exactitud. Esta es la segunda razón de que los polinomios algebraicos constituyan un medio universal de aproximar funciones. La mera sencillez de construcción de polinomios algebraicos no es suficiente; necesitamos también la posibilidad de aproximar una función continua por medio de un polinomio con un error arbitrariamente prefijado. Esta posibilidad fue demostrada por Weierstrass.

Las profundas ideas de Chebichev sobre la aproximación óptima y el teorema de Weierstrass sirvieron de base, a principios de este siglo, para el desarrollo actual de la teoría de aproximación. En relación con esto mencionemos los nombres de S. N. Bernstein, Borel, Jackson, Lebesgue y la Vallée-Poussin. Este desarrollo se puede describir brevemente como sigue. Hasta Chebichev (principios de este siglo), los problemas consistían generalmente en la aproximación de funciones concretas, mientras que el problema característico del momento actual es la aproximación, por polinomios o de otro modo, de clases enteras de funciones, analíticas, diferenciables, etc.

La escuela rusa, y ahora la escuela soviética, de la teoría de aproximación de funciones ha jugado un papel capital en dicha teoría. S. N. Bernstein, A. N. Kolmogorov, M. A. Laurentiev y sus discípulos son autores de importantes contribuciones. Actualmente la teoría se ha desarrollado como una rama esencialmente distinta de la teoría de funciones.

Además de los polinomios algebraicos, otro medio muy importante de aproximación consiste en los polinomios trigonométricos. Un *polinomio trigonométrico* de orden n es una función de la forma

$$u_n(x) = a_0 + a_1 \cos x + \beta_1 \sin x + a_2 \cos 2x + \beta_2 \sin 2x + \\ + \cdots + a_n \cos nx + \beta_n \sin nx,$$

o más brevemente

$$u_n(x) = a_0 + \sum_{k=1}^n (a_k \cos kx + \beta_k \operatorname{sen} kx),$$

donde a_k y β_k son constantes.

Existen distintos métodos particulares de aproximación mediante polinomios trigonométricos, generalmente relacionados de una forma más bien simple con los métodos correspondientes de polinomios algebraicos. Entre ellos juega un papel especialmente importante el desarrollo de funciones en serie de Fourier (véase §7). Estas series se conocen con el nombre del matemático francés Fourier, quien al principio del siglo pasado realizó varios descubrimientos teóricos relacionados con ellas en su estudio de la propagación del calor. Sin embargo, debemos observar que las series trigonométricas fueron estudiadas ya hacia mediados del siglo XVIII por los grandes matemáticos Leonhard Euler y Daniel Bernoulli. En el trabajo de Euler estaban relacionadas con sus investigaciones de astronomía, y en el de Bernoulli, con su estudio de la cuerda vibrante. Señalemos que Euler y Bernoulli suscitaron la cuestión particularmente importante de la posibilidad de representar una función más o menos arbitraria mediante una serie trigonométrica, cuestión que no quedó definitivamente resuelta hasta mediados del siglo pasado. La respuesta (afirmativa, y que discutiremos más adelante) fue anticipada por Bernoulli.

Las series de Fourier son de gran importancia en física, pero nosotros prestaremos poca atención a este aspecto, ya que ha sido considerado ya en el capítulo 6. En dicho capítulo, el lector también encontrará ejemplos de problemas físicos que llevaron naturalmente al desarrollo de una función en series distintas de las trigonométricas pero muy similares. Nos referimos a las llamadas series de funciones ortogonales.

Las series de Fourier tienen una historia de doscientos años, por lo cual no es sorprendente que actualmente esta teoría sea extraordinariamente amplia, sutil y profunda y constituya una disciplina matemática independiente. En esta teoría ha representado un papel especialmente notable la escuela de Moscú de la teoría de funciones de una variable real, N. N. Luzin, A. N. Kolmogorov, D. E. Menshov y otros.

Observemos también que la importancia de los polinomios trigonométricos en la matemática moderna apenas se agota en su papel como método de aproximación. Por ejemplo, en el capítulo 10 el

lector llegó a familiarizarse con los resultados fundamentales de I. M. Vinogradov en la teoría de números, derivados de un aparato convenientemente construido de sumas trigonométricas (polinomios).

§2. Polinomios de interpolación

Un caso especial de construcción de polinomios de interpolación

En la práctica se utiliza muy a menudo el método de interpolación

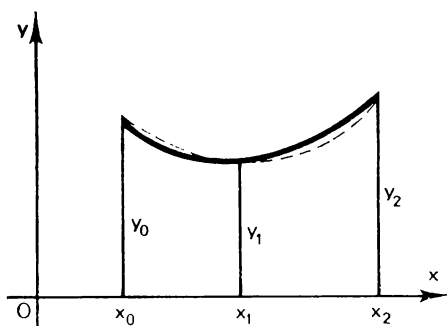


FIG. 1

para aproximar una función. A fin de introducir al lector en cuestiones de este tipo, consideremos el siguiente problema elemental.

Sea la función $y = f(x)$ definida en el intervalo $[x_0, x_2]$, cuya gráfica se ve en la figura 1. El aspecto de esta gráfica recuerda a un arco de parábola. Así, si deseamos aproximar nuestra función mediante otra sencilla, es natural escoger un polinomio de segundo grado

$$P(x) = a_0 + a_1x + a_2x^2, \quad (1)$$

cuya gráfica es una parábola.

El método de interpolación consiste en lo siguiente. En el intervalo $[x_0, x_2]$ elegimos un punto interior x_1 . Los puntos x_0, x_1, x_2 dan valores correspondientes de nuestra función

$$y_0 = f(x_0), \quad y_1 = f(x_1), \quad y_2 = f(x_2).$$

Construimos un polinomio (1) tal que en los puntos x_0, x_1, x_2 coincida con la función en cuestión (su gráfica se indica por la línea de puntos en la figura 1). En otras palabras, debemos elegir los coeficientes a_0, a_1, a_2 en el polinomio (1) de modo tal que verifiquen las ecuaciones

$$P(x_0) = y_0, \quad P(x_1) = y_1, \quad P(x_2) = y_2. \quad (2)$$

Observemos que nuestra función puede no venir definida por una fórmula; por ejemplo, sus valores pueden estar dados empíricamente tal como se indica en la gráfica de la figura 1. Para resolver el problema de interpolación, elegimos una función aproximadora en forma de una expresión analítica: el polinomio $P(x)$. Si la exactitud de la aproximación es satisfactoria, el polinomio elegido de esta forma tiene la ventaja, sobre la función original, de que podemos calcular sus valores intermedios.

Este problema de interpolación podría resolverse como sigue: escribamos las tres ecuaciones:

$$y_0 = a_0 + a_1x_0 + a_2x_0^2,$$

$$y_1 = a_0 + a_1x_1 + a_2x_1^2,$$

$$y_2 = a_0 + a_1x_2 + a_2x_2^2,$$

resolvámoslas para a_0, a_1, a_2 y sustituyamos los valores de estos coeficientes en la ecuación (1). Pero resolvámoslas de una forma algo diferente. Empecemos por construir el polinomio $Q_0(x)$ de segundo grado que verifique las tres condiciones: $Q_0(x_0) = 1$, $Q_0(x_1) = 0$, $Q_0(x_2) = 0$. De las dos últimas condiciones se deduce que este polinomio debe ser de la forma $A(x - x_1)(x - x_2)$, y de la primera condición, que

$$A = \frac{1}{(x_0 - x_1)(x_0 - x_2)}.$$

Así, el polinomio buscado tiene la forma

$$Q_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}.$$

Análogamente, los polinomios

$$Q_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}, \quad Q_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

verifican las condiciones

$$\begin{aligned} Q_1(x_0) = Q_1(x_2) = 0, & \quad Q_1(x_1) = 1, \\ Q_2(x_0) = Q_2(x_1) = 0, & \quad Q_2(x_2) = 1. \end{aligned}$$

Además, es evidente que el polinomio $y_0Q_0(x)$ toma el valor y_0 para $x = x_0$ y se anula para $x = x_1$ y $x = x_2$, cumpliéndose propiedades correspondientes para los polinomios $y_1Q_1(x)$ e $y_2Q_2(x)$.

De aquí se deduce fácilmente que el polinomio interpolador buscado está dado por la fórmula

$$\begin{aligned} P(x) &= y_0Q_0(x) + y_1Q_1(x) + y_2Q_2(x) = \\ &= y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + \\ &\quad + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}. \end{aligned} \quad (3)$$

Observemos que el polinomio así obtenido es el único de segundo grado que resuelve nuestro problema de interpolación. Pues, si suponemos que otro polinomio de segundo grado $P_1(x)$ es también solución del problema, entonces la diferencia $P_1(x) - P(x)$, que es de segundo grado, se anula en los tres puntos $x = x_0, x_1, x_2$. Pero sabemos por el álgebra que, si un polinomio de segundo grado se anula para tres valores de x , entonces es idénticamente nulo. Por tanto, los polinomios $P(x)$ y $P_1(x)$ son idénticos.

Está claro que en general el polinomio así obtenido coincide con la función dada solamente en los puntos x_0, x_1, x_2 y es distinto para otros valores de x .

Si tomamos x_1 en el centro del intervalo $[x_0, x_2]$ y ponemos $x_2 - x_1 = x_1 - x_0 = h$, entonces la fórmula (3) se simplifica algo:

$$\begin{aligned} P(x) &= \frac{1}{2h^2} [y_0(x-x_1)(x-x_2) - 2y_1(x-x_0)(x-x_2) + \\ &\quad + y_2(x-x_0)(x-x_1)]. \end{aligned}$$

Como ejemplo vamos a interpolar la curva $y = \sin x$ (Fig. 2) mediante un polinomio de grado dos que coincida con ella en los puntos $x = 0, \pi/2, \pi$. Evidentemente, el polinomio tiene la forma

$$P(x) = \frac{4}{\pi^2} x(\pi - x) \approx \sin x.$$

Comparemos $\sin x$ y $P(x)$ en dos puntos intermedios:

$$P\left(\frac{\pi}{4}\right) = 0,75, \text{ mientras } \sin \frac{\pi}{4} = \frac{\sqrt{2}}{2} \approx 0,71.$$

$$P\left(\frac{\pi}{6}\right) = \frac{10}{18}, \text{ mientras } \sin \frac{\pi}{6} = \frac{1}{2}.$$

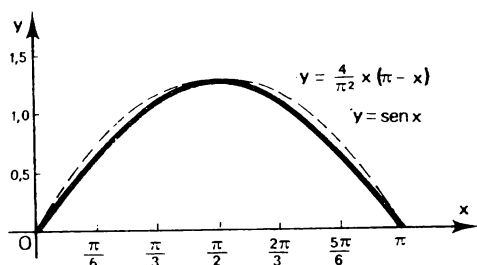


FIG. 2

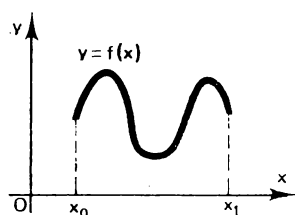


FIG. 3

De este modo hemos aproximado $\sin x$ en el intervalo $[0, \pi]$ con un error¹ de un 0,05. Por otra parte, el desarrollo de $\sin x$ en serie de Taylor en torno al punto $\pi/2$ da

$$\sin x = \cos\left(\frac{\pi}{2} - x\right) = 1 - \frac{\left(\frac{\pi}{2} - x\right)^2}{2!} + \frac{\left(\frac{\pi}{2} - x\right)^4}{4!} - \dots$$

Si nos detenemos en el segundo término del desarrollo, tenemos en el punto $x = 0$ la aproximación $\sin 0 = 1 - \pi^2/8 \approx 0,234$, es decir, un error mayor que 0,2.

Vemos que nuestro método de interpolación ha conseguido una aproximación (de $\sin x$ sobre el intervalo $[0, \pi]$ mediante un polinomio de grado dos) más satisfactoria que el desarrollo de Taylor de segundo grado. Sin embargo, no debemos olvidar que la fórmula

¹ Sin embargo, para una completa justificación de esta afirmación, necesitamos probar que la diferencia $(4x/\pi^2)(\pi - x) - \sin x$ no supera el valor 0,05, no sólo para $x = \pi/4$ y $x = \pi/6$, sino para todo x en el intervalo $[0, \pi]$; no lo haremos.

de Taylor en torno al punto $x = \pi/2$ da una aproximación más exacta que la aproximación obtenida por interpolación.

Solución general del problema

Está claro que una función más complicada $y = f(x)$, como la de la figura 3, no se presta a una aproximación mediante un polinomio de grado dos, ya que ninguna parábola de segundo grado se ajustaría a tal función $y = f(x)$. En este caso es natural intentar una interpolación con un polinomio de grado superior (no menor que cuatro).

El problema general de interpolación consiste en construir un polinomio $P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ de grado n que coincida con la función dada en $n + 1$ puntos:

$$P(x_0) = f(x_0), P(x_1) = f(x_1), \dots, P(x_n) = f(x_n).$$

Los puntos en los que se exige que la función coincida con su polinomio aproximador se denominan *puntos de interpolación*.

Razonando de la misma forma que para un polinomio de segundo grado, podemos probar fácilmente que el polinomio buscado se puede escribir en la forma

$$P_n(x) = \sum_{k=0}^n \frac{(x - x_0)(x - x_1) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0)(x_k - x_1) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)} f(x_k), \quad (4)$$

y además que este polinomio (de grado n) es único. La fórmula así escrita se conoce como *fórmula de Lagrange*. También se puede poner en otras formas distintas; por ejemplo, en la práctica se utiliza mucho escrita en términos de cocientes de diferencias de Newton.

Error de interpolación

El método de interpolación es un medio universal de aproximar funciones. En principio no es necesario que la función tenga propiedades especiales para que la interpolación sea posible; por ejemplo no es necesario que tenga derivadas sobre el intervalo total de aproximación. En este aspecto el método de interpolación aventaja a la fórmula de Taylor. Es interesante observar que hay casos en que la función es incluso analítica en cada punto de un intervalo y en cambio no se puede aproximar por la fórmula de Taylor sobre el intervalo. Supongamos, por ejemplo, que necesitamos una buena aproxima-

ción de la función $1/(1+x^2)$ sobre el intervalo $[-2, 2]$ mediante un polinomio algebraico. De entrada es natural intentar su desarrollo en serie de Taylor en torno al punto $x=0$

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \dots$$

Pero es fácil ver que esta serie solamente converge en el intervalo $-1 < x < 1$. Fuera del intervalo $(-1, 1)$ diverge, y por consiguiente no puede aproximar a $1/(1+x^2)$ sobre todo el intervalo $[-2, 2]$. No obstante, el método de interpolación es perfectamente aplicable aquí.

Naturalmente, en cada caso particular surge la cuestión de elegir el número y distribución de los puntos de interpolación de tal forma que el error se ajuste a ciertas exigencias. Para funciones con derivadas de orden suficientemente alto, la respuesta a la cuestión de la posible magnitud del error viene dada por el resultado clásico siguiente, que introducimos sin demostración.

Si la función $f(x)$ tiene derivada continua de orden $n+1$ sobre el intervalo $[x_0, x_n]$, entonces para cualquier valor intermedio x la desviación de $f(x)$ respecto del polinomio de interpolación de Lagrange $P(x)$ con puntos de interpolación $x_0 < x_1 < \dots < x_n$ está dada por la fórmula

$$f(x) - P(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_n)}{n!} f^{(n+1)}(c),$$

donde c es un punto intermedio entre x_0 y x_n . Esta fórmula recuerda a la del resto del desarrollo de Taylor y es esencialmente una generalización de ésta. Así, si sabemos que la derivada $f^{(n+1)}(x)$ de orden $n+1$ no excede del número M en valor absoluto sobre el intervalo $[x_0, x_n]$, entonces el error de la aproximación para cualquier valor de x sobre este intervalo está acotado del siguiente modo:

$$|f(x) - P_n(x)| \leq \frac{|x-x_0| \dots |x-x_n|}{n!} M.$$

La moderna teoría de aproximación proporciona otros muchos métodos de estimar el error de interpolación. Esta cuestión ha sido estudiada cuidadosamente, descubriendo algunos hechos interesantes completamente inesperados.

Consideremos, por ejemplo, una función regular $y = f(x)$, definida sobre el intervalo $[-1, 1]$, tal que su gráfica sea una curva continua cuya tangente varíe continuamente. Nuestra elección del intervalo con extremos concretos -1 y 1 carece de importancia; el hecho considerado aquí sigue siendo válido para un intervalo arbitrario $[a, b]$, sin más que algunos cambios irrelevantes.

Supongamos ahora que sobre el intervalo $[-1, 1]$ elegimos un sistema de $n + 1$ puntos

$$-1 \leq x_0 < x_1 < \cdots < x_n \leq 1 \quad (5)$$

y a continuación construimos el polinomio $P(x) = a_0 + a_1x + \cdots + a_nx^n$ de grado n que coincida con $f(x)$ en estos puntos. Supondremos por el momento que los puntos del sistema (5) están uniformemente espaciados a lo largo del intervalo. Si n crece indefinidamente, entonces el correspondiente polinomio interpolador $P_n(x)$ coincidirá con $f(x)$ en un número de puntos cada vez mayor, y podríamos pensar que en un punto intermedio x , no perteneciente al sistema (5), la diferencia $f(x) - P_n(x)$ tendería a cero cuando $n \rightarrow \infty$. Esta opinión se mantuvo hasta finales del siglo pasado, pero después se descubrió que la realidad es otra. Se ha probado que para muchas funciones regulares (incluso analíticas) $f(x)$, en el caso de puntos espaciados regularmente x_k , los polinomios interpoladores $P_n(x)$ no convergen en absoluto a $f(x)$ cuando $n \rightarrow \infty$. La gráfica del polinomio interpolador coincide indudablemente con $f(x)$ en los puntos de interpolación dados, mas para n grande se desvía fuertemente de la gráfica de $f(x)$ en valores intermedios de x , y la desviación aumenta al crecer n . Como demostraron ulteriores investigaciones, esta situación se puede evitar, al menos para funciones regulares, si los puntos de interpolación se distribuyen de un modo más disperso cerca del centro del intervalo y más densamente cerca de los extremos. En realidad, se ha demostrado que, en un sentido bien conocido, la distribución óptima de los puntos de interpolación es aquella en que los puntos x_k se hallan en los ceros² de los polinomios de Chebichev $\cos(n + 1) \arccos x$ definidos por la fórmula

$$x_k = \cos \frac{2k + 1}{2(n + 1)} \pi \quad (k = 0, 1, \cdots, n).$$

² Un cero de la función $f(x)$ es un valor x_k para el que $f(x_k) = 0$. Para más detalles sobre polinomios de Chebichev, véase §5.

Los polinomios (llamados de Chebichev) que corresponden a estos puntos de interpolación tiene la propiedad de converger uniformemente a la función que los engendra, siempre que ésta sea regular, esto es, sea continua y tenga derivada primera continua. La gráfica de tal función es una curva continua cuya tangente varía continuamente. La figura 4 muestra la distribución de los ceros del polinomio de Chebichev para el caso $n = 5$.

En cuanto a funciones continuas no regulares arbitrarias la situación es peor; se puede probar que en general no existe ninguna sucesión de puntos de interpolación tal que el proceso de interpolación converja para cualquier función continua (teorema de Faber). En otras palabras, como quiera que dividamos el intervalo $[-1, 1]$, con el número de puntos de interpolación tendiendo a infinito, podemos siempre hallar una función $f(x)$, continua en el intervalo, tal que los polinomios sucesivos (cor-

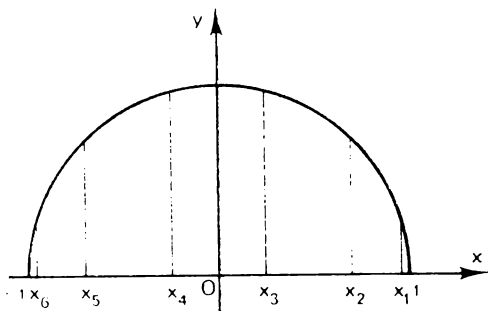


FIG. 4

respondientes a estos puntos de interpolación) no converjan a la función. Incluso los matemáticos de mediados del siglo pasado, de haber conocido este hecho, lo habrían calificado de paradójico. Naturalmente, la explicación es que entre las funciones continuas no regulares hay algunas extraordinariamente «malas», como son, por ejemplo, aquellas que no tienen derivada en un punto del intervalo en que están definidas; estas funciones suministran ejemplos para los que un proceso de interpolación dado no convergerá. Existen algunos métodos de aproximación para estas funciones, basados en ciertas modificaciones del proceso de interpolación anterior, aunque no nos detendremos en ello.

Para terminar observemos que los polinomios algebraicos no son los únicos medios disponibles de interpolación. Existen métodos de polinomios trigonométricos, por ejemplo, que están bien desarrollados desde el punto de vista práctico e incluso teórico.

§3. Aproximación de integrales definidas

La interpolación de funciones tiene amplias aplicaciones en cuestiones relativas al cálculo aproximado de integrales. Como ejemplo, introducimos una fórmula aproximada para la integral definida, la regla de Simpson, que es muy utilizada en análisis aplicado.

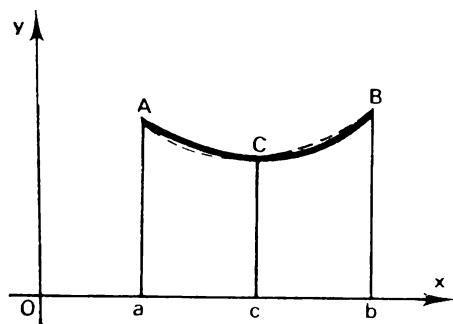


FIG. 5

Deseamos calcular una aproximación de la integral de la función $f(x)$ definida sobre el intervalo $[a, b]$, cuya gráfica se ve en la figura 5. El valor exacto viene dado por el área del trapecio curvilíneo $aABb$. Sea C el punto de la gráfica abscisa $c = (a + b)/2$. Por los puntos A, B y C hacemos pasar una parábola de

segundo grado. Como sabemos por la sección anterior, esta parábola es la gráfica de un polinomio de segundo grado, definido por la ecuación

$$P(x) = \frac{1}{2h^2} [(x - c)(x - b)y_0 - 2(x - a)(x - b)y_1 + (x - a)(x - c)y_2],$$

donde

$$h = \frac{b - a}{2}, \quad y_0 = f(a), \quad y_1 = f(c), \quad y_2 = f(b).$$

Con la terminología de la sección precedente, podemos decir que el polinomio de segundo grado $P(x)$ interpola a $f(x)$ en los puntos de abscisa a, c, b . Si la gráfica de la función $f(x)$ sobre el intervalo $[a, b]$ no cambia demasiado bruscamente y el intervalo no es grande, entonces $P(x)$ diferirá poco de $f(x)$ en todos los puntos; esto implica que sus integrales sobre $[a, b]$ diferirán poco una de otra. Sobre esta

base podemos suponer que estas integrales son aproximadamente iguales,

$$\int_a^b f(x) dx \approx \int_a^b P(x) dx,$$

o, como suele decirse, la segunda integral es una aproximación de la primera. Cálculos sencillos, que dejamos al lector, muestran que

$$\int_a^b (x-c)(x-b) dx = \frac{2}{3} h^3, \quad -\int_a^b (x-a)(x-b) dx = \frac{4}{3} h^3,$$

$$\int_a^b (x-a)(x-c) dx = \frac{2}{3} h^3.$$

De aquí

$$\int_a^b P(x) dx = \frac{h}{3} [f(a) + 4f(c) + f(b)].$$

Así, la integral definida se puede calcular por la siguiente fórmula de aproximación

$$\int_a^b f(x) dx \approx \frac{h}{3} [f(a) + 4f(c) + f(b)].$$

Esta es la fórmula de Simpson.

Como ejemplo, utilicemos esta fórmula para calcular la integral de $\sin x$ en el intervalo $[0, \pi]$. En este caso

$$h = \frac{\pi}{2}, \quad f(a) = \sin 0 = 0, \quad f(c) = \sin \frac{\pi}{2} = 1, \quad f(b) = \sin \pi = 0,$$

y por consiguiente $(h/3)[f(a) + 4f(c) + f(b)] = (2/3)\pi = 2,09\dots$. Por otra parte, la integral se puede calcular exactamente

$$\int_0^{\pi} \sin x dx = -\cos x \Big|_0^{\pi} = 2.$$

El error no excede de 0,1.

Si el intervalo $[0, \pi]$ se descompone en dos partes iguales y aplicamos nuestra fórmula sobre cada una de ellas por separado, entonces obtenemos

$$\int_0^{\pi/2} \operatorname{sen} x \, dx \approx \frac{\pi}{12} \left[\operatorname{sen} 0 + 4 \operatorname{sen} \frac{\pi}{4} + \operatorname{sen} \frac{\pi}{2} \right] = \frac{\pi}{12} \left(4 \frac{\sqrt{2}}{2} + 1 \right) \approx 1,001,$$

$$\int_{\pi/2}^{\pi} \operatorname{sen} x \, dx \approx 1,001.$$

De esta manera

$$\int_0^{\pi} \operatorname{sen} x \, dx \approx 2,002;$$

y ahora el error es considerablemente menor que 0,002.

En la práctica, para calcular aproximadamente la integral definida de una función $f(x)$ sobre $[a, b]$ dividimos el intervalo en un número par de partes mediante los puntos $a = x_0 < x_1 < \dots < x_n = b$ y aplicamos la regla de Simpson al segmento $[x_0, x_2]$, después al $[x_2, x_4]$ y así sucesivamente. Como resultado tenemos la siguiente fórmula general de Simpson:

$$\int_a^b f(x) \, dx \approx \frac{b-a}{3n} [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots + f(x_n)]. \quad (6)$$

Demos ahora sin demostración la estimación clásica del error. Si la función $f(x)$ tiene derivada cuarta en el intervalo $[a, b]$ y ésta verifica la desigualdad $|f^{IV}(x)| \leq M$, entonces la siguiente estimación es válida

$$\left| \int_a^b f(x) \, dx - L(f) \right| \leq \frac{M(b-a)^5}{180n^4}, \quad (7)$$

donde $L(f)$ representa el segundo miembro de la fórmula (6). En este caso el error será de orden n^{-4} ³.

Podríamos haber descompuesto el intervalo $[a, b]$ en n partes iguales y tomado como aproximación de la integral la suma de las

³ Si una cierta cantidad α_n , que depende de $n = 1, 2, \dots$, verifica la desigualdad $|\alpha_n| < C/n^k$, donde C es constante independiente de n , entonces decimos que es de orden n^{-k} .

áreas de los rectángulos dibujados en la figura 6. Entonces obtendríamos una fórmula de aproximación por rectángulos ⁴

$$\int_a^b f(x) dx \approx \frac{b-a}{n} [f(x_0) + f(x_1) + \cdots + f(x_{n-1})]. \quad (8)$$

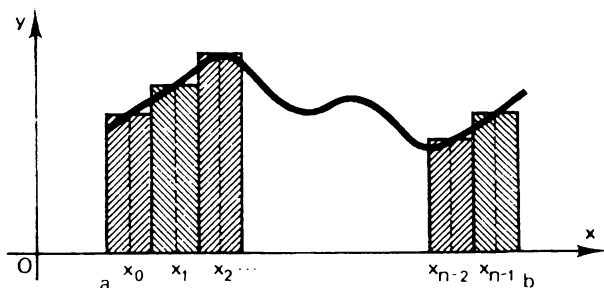


FIG. 6

Se puede probar que el orden del error es aquí n^{-2} , suponiendo que la función tenga derivada segunda acotada en el intervalo $[a, b]$. Podemos también tomar como aproximación la suma de las áreas de los trapecoides dibujados en la figura 7, obteniendo la fórmula trapezoidal

$$\int_a^b f(x) dx \approx \frac{b-a}{2n} [f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)] \quad (9)$$

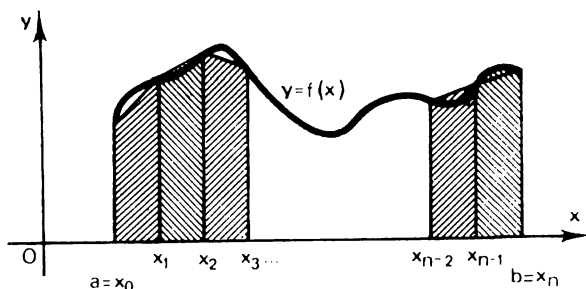


FIG. 7

⁴ En este caso $x_0, x_1, \dots, x_{n-1}$ son los centros de las partes iguales del intervalo $[a, b]$, y no puntos de división como en las fórmulas (6) y (9).

con orden de error n^{-2} , suponiendo que la función tenga derivada segunda acotada.

Se suele afirmar que la fórmula de Simpson es más exacta que la rectangular y trapezoidal. Esta afirmación es preciso matizarla, pues de otro modo no es cierta. Si sabemos solamente que la función tiene derivada primera, entonces el orden de error garantizado para cada uno de los tres métodos es igual a n^{-1} ; en este caso la fórmula de Simpson no tiene una ventaja esencial sobre la rectangular y trapezoidal. Para funciones con derivada segunda, está garantizada una aproximación de orden n^{-2} tanto para la fórmula trapezoidal como para la de Simpson. Pero si la función tiene derivadas tercera y cuarta, entonces el orden de error será igual a n^{-2} para las fórmulas trapezoidal y rectangular, pero para la de Simpson es igual a n^{-3} y n^{-4} , respectivamente. Ahora bien, el orden n^{-4} resulta ser el mejor resultado posible para la fórmula de Simpson; en otras palabras, para funciones con derivadas de orden superior al cuarto, el orden de error sigue siendo n^{-4} . Así, si tenemos una función con derivada quinta y queremos aprovechar este hecho para obtener una aproximación de orden n^{-5} , necesitamos un nuevo método de aproximación para integrales definidas, distinto de la fórmula de Simpson. Para exponer cómo construirlo, observemos lo siguiente.

Las fórmulas trapezoidal y rectangular, como se prueba fácilmente, son exactas para polinomios de primer grado; esto significa que escribiendo en (9) la función $A + Bx$, donde A y B son constantes, llegamos a una igualdad exacta. Del mismo modo, la fórmula de Simpson resulta ser exacta para polinomios de tercer grado $A + Bx + Cx^2 + Dx^3$. Este hecho es el quid de la cuestión. Supongamos que hemos dividido el intervalo $[a, b]$ en n partes iguales y que sobre cada una hemos utilizado un método de aproximación —el mismo en cada una de ellas— que es exacto para polinomios $A + Bx + \cdots + Fx^{m-1}$ de grado $m - 1$. Entonces el error de aproximación para toda función con derivada m -ésima acotada será de orden n^{-m} , y si la función no es un polinomio de grado $m - 1$, entonces este orden no se puede aumentar ni siquiera para funciones con derivadas de orden superior.

Estas observaciones subrayan la importancia de encontrar el método aproximado de integración más sencillo posible que sea exacto para polinomios de un grado determinado. Esta cuestión, sobre la que actualmente existe mucha literatura, ha interesado a los matemáticos desde hace mucho. Aquí solamente podemos hacer referencia a algunos resultados clásicos.

Sea una función $p(x)$ dada. Nos preguntamos cómo distribuir sobre el intervalo $[-1, 1]$ los puntos de división $x_1, \dots, x_m$ y elegir el número K , para que se verifique la ecuación

$$\int_{-1}^1 f(x) p(x) dx = K \sum_{i=1}^m f(x_i),$$

para todo polinomio $f(x)$ de grado m .

Se puede demostrar que para $p(x) = (1 - x^2)^{-1/2}$ el problema está resuelto si $K = \pi/m$, y x_i son los ceros del polinomio de Chebichev $\cos m \arccos x$ (véase §5).

Para $p(x) = 1$, Chebichev dio una solución del problema para $m = 1, 2, \dots, 7$. Para $m = 8$ el problema no tiene solución: los puntos de división se pueden hallar, pero son complejos. Para $m = 9$ vuelve a tener solución. Sin embargo, como demostró S. N. Bernstein, para $m > 9$ el problema no tiene solución (los puntos de división caen fuera del intervalo $[-1, +1]$).

Haciendo uso de la fórmula de Lagrange (4) se puede construir muy fácilmente una fórmula de cuadratura exacta para polinomios de grado n . Si la integramos miembro a miembro sobre el intervalo $[a, b]$, obtenemos

$$\int_a^b P_n(x) dx = \sum_{k=0}^n p_k f(x_k), \quad (10)$$

donde

$$p_k = \int_a^b \frac{(x - x_0) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)} dx$$

$$(k = 0, 1, \dots, n).$$

Por consiguiente, la ecuación (10) es válida para todos los polinomios de grado n , y así la fórmula de cuadratura

$$\int_a^b f(x) dx \approx \sum_{k=0}^n p_k f(x_k)$$

es exacta para todos los polinomios de grado n .

Cuando

$$x_0 = a, \quad x_1 = \frac{a+b}{2}, \quad x_2 = b,$$

esta fórmula se reduce, como hemos visto, a la de Simpson.

La distribución de los puntos de interpolación x_k ($k = 0, 1, \dots, n$) en el intervalo $[a, b]$ se puede cambiar. Para cada distribución de los puntos habrá una fórmula de cuadratura correspondiente.

Gauss, el famoso matemático alemán del siglo pasado, demostró que los puntos de interpolación x_k se pueden distribuir de tal manera que la fórmula sea exacta para todos los polinomios no sólo de grado n , sino también de grado $2n + 1$.

El polinomio

$$A_{n+1}(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$$

de grado $n + 1$, construido con los puntos de división de Gauss x_k , tiene una notable propiedad: para cualquier polinomio $P(x)$ de grado menor que $n + 1$, tenemos la ecuación

$$\int_a^b A_{n+1}(x) P(x) dx = 0.$$

Dicho con otras palabras, el polinomio $A_{n+1}(x)$ es ortogonal, sobre el intervalo $[a, b]$, a todos aquellos cuyo grado no sea superior a n . Los $A_{n+1}(x)$ se denominan *polinomios de Legendre* (correspondientes al intervalo $[a, b]$).

§4. Concepto de aproximación uniforme óptima de Chebichev

Planteamiento del problema

Chebichev llegó a la idea de aproximación uniforme óptima a través de un problema puramente práctico, ya que, además de uno de los más grandes matemáticos del siglo pasado y creador de la base de varias disciplinas matemáticas extensamente desarrolladas en la actualidad, fue también un ingeniero relevante de su tiempo. En particular, Chebichev estuvo muy interesado en la construcción de mecanismos productores de una trayectoria de movimiento dada. A continuación explicaremos esta idea.

Sea la curva $y = f(x)$ sobre el intervalo $a \leq x \leq b$. Deseamos construir, bajo determinados requisitos técnicos, un mecanismo tal que uno de sus puntos describa esta curva tan exactamente como sea posible al ponerlo en funcionamiento. Chebichev resolvió el problema como sigue. En primer lugar, buscando la solución como ingeniero,

construyó el mecanismo requerido de suerte que se obtuviera una aproximación cruda de la trayectoria pedida. Así, un cierto punto A del mecanismo (en su forma provisional) describiría la curva

$$y = \phi(x), \quad (11)$$

parecida a la curva exigida $y = f(x)$ sólo en su aspecto general. El mecanismo así construido consta de piezas, engranajes, palancas de varias clases y similares. Todas ellas tienen medidas específicas

$$a_0, a_1, a_2, \dots, a_m, \quad (12)$$

que describen completamente el mecanismo, y por consiguiente la curva (11). Son los parámetros del mecanismo y de la curva (11)⁵. Así, la curva (11) depende no sólo de la variable x , sino también de los parámetros (12). A un sistema de valores de los parámetros corresponderá una curva específica, cuya ecuación se puede escribir en la forma

$$y = \phi(x; a_0, a_1, \dots, a_m). \quad (13)$$

En tales casos se suele decir que hemos obtenido una familia de funciones (13), definida sobre el intervalo $a \leq x \leq b$ y dependiente de los $m + 1$ parámetros (12).

En los restantes pasos de la resolución del problema, Chebichev trabajó como matemático puro. Propuso, de una manera perfectamente natural, tomar como medida de la desviación de la función $f(x)$ respecto de la función aproximadora $\phi(x; a_0, a_1, \dots, a_m)$ la magnitud

$$||f - \phi|| = \max_{a \leq x \leq b} |f(x) - \phi(x; a_0, a_1, \dots, a_m)|, \quad (14)$$

igual al máximo del valor absoluto de la diferencia $f(x) - \phi(x; a_0, a_1, \dots, a_m)$ sobre el intervalo $a \leq x \leq b$ (Fig. 8). Esta cantidad es evidentemente una cierta función

$$||f - \phi|| = F(a_0, a_1, \dots, a_m) \quad (15)$$

de los parámetros $a_0, a_1, \dots, a_m$. El problema consiste ahora en hallar aquellos valores de los parámetros para los que la función (15)

⁵ Los detalles de los cálculos para mecanismos de este tipo se pueden encontrar en la publicación «The Scientific Heritage of P. L. Chebichev», volumen II, Academia de Ciencias de la URSS, 1945.

es mínima. Estos valores definen una función ϕ , que se acostumbra a describir como la aproximación uniforme óptima de la función dada $y = f(x)$ entre todas las funciones posibles de la familia dada (11). La magnitud $F(\alpha_0, \alpha_1, \dots, \alpha_m)$ para estos valores de los parámetros se llama la *aproximación uniforme óptima de la función $f(x)$ sobre el intervalo $[a, b]$ mediante las funciones de la familia (13)*. Se

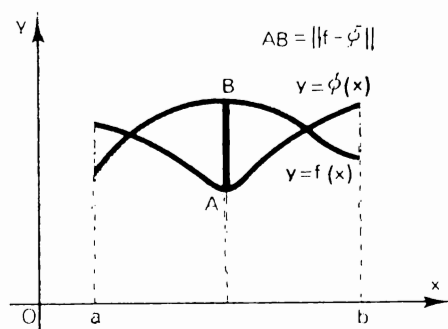


FIG. 8

suele denotar con el símbolo $E_m(f)$. El término «uniforme» se sustituye a menudo, especialmente en la literatura no soviética, por el término «Chebichev». Ambos destacan el carácter especial de la aproximación, ya que, naturalmente, son posibles otros tipos de aproximación; por ejemplo, se puede hablar de la aproximación óptima de $f(x)$ por funciones de una

familia dada en el sentido del cuadrado medio. Esto se discutirá en el §8.

Chebichev descubrió en primer lugar las distintas leyes que valen para el tipo de aproximación de que se trata aquí y encontró que en muchos casos la función ϕ , que es la aproximación uniforme óptima de $f(x)$ sobre el intervalo $[a, b]$, tiene la notable propiedad de que para ella el máximo (15) del valor absoluto de la diferencia

$$f(x) - \phi(x; \alpha_0, \alpha_1, \dots, \alpha_m)$$

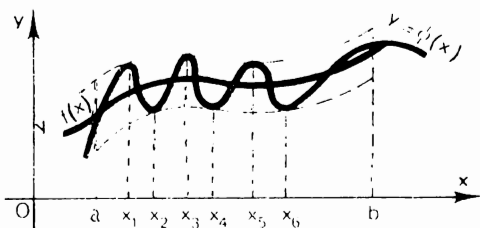


FIG. 9

se alcanza para $m + 2$ puntos al menos del intervalo $[a, b]$ con signos sucesivos alternados (Figura. 9).

No disponemos de espacio aquí para una formulación exacta de las condiciones bajo las cuales es válida esta proposición, por lo que remitimos al lector al artículo de V. L. Goncharov «The theory of the best approximation of functions» («La teoría de la aproximación

óptima de funciones») (*The Scientific Heritage of Chebichev*, volumen I).

El caso de aproximación de funciones por polinomios

Las investigaciones de Chebichev antes citadas son especialmente importantes para la teoría general de aproximación cuando se aplican al problema de aproximar una función arbitraria $f(x)$ sobre un intervalo $[a, b]$ por polinomios $P_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ de grado n dado. Los polinomios $P_n(x)$ de grado n constituyen una familia de funciones dependientes de los $n + 1$ coeficientes como parámetros. Como se puede demostrar, la teoría de Chebichev es totalmente aplicable a polinomios, por lo que si queremos efectuar la aproximación uniforme óptima de la función $f(x)$ en el segmento $[a, b]$ por un polinomio $P_n(x)$ elegido entre todos los

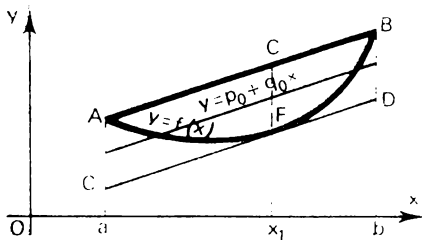


FIG. 10

polinomios posibles de grado n dado, solamente necesitamos hallar todos aquellos valores de x de este intervalo para los cuales la función $|f(x) - P_n(x)|$ alcanza su máximo L sobre $[a, b]$. Si entre ellos podemos encontrar $n + 2$ valores $x_1, x_2, \dots, x_{n+2}$, tales que la diferencia $f(x) - P_n(x)$ cambie de signo sucesivamente

$$f(x_1) - P_n(x_1) = \pm L,$$

$$f(x_2) - P_n(x_2) = \pm L,$$

$$\dots\dots\dots$$

$$f(x_{n+2}) - P_n(x_{n+2}) = \pm (-1)^{n+1} L,$$

entonces $P_n(x)$ es el polinomio óptimo, y en caso contrario no. Por ejemplo, la solución del problema de la aproximación uniforme óptima de la función $f(x)$, ilustrada en la figura 10, mediante polinomios $P_1(x) = p + qx$ de primer grado consiste en el polinomio $p_0 + q_0x$, cuya gráfica es una recta paralela a la cuerda AB y que divide en partes iguales al paralelogramo comprendido entre la cuerda y la tangente CD paralela a la cuerda; pues el valor absoluto

de la diferencia $f(x) - (p_0 + q_0x)$ evidentemente alcanza su máximo para los valores $x_0 = a$, x_1 y $x_2 = b$, donde x_1 es la abscisa del punto de tangencia F , y para estos valores la diferencia cambia sucesivamente de signo. Para evitar interpretaciones erróneas, observemos que estamos hablando de una curva cóncava hacia arriba con tangente en todo punto. En este ejemplo $E_1(f)$ es igual a la mitad de la longitud de cualquiera de los segmentos (iguales) AC , BD o GF .

§5. Los polinomios de Chebichev cuya desviación de cero es mínima

Consideremos el siguiente problema. Queremos hallar un polinomio $P_{n-1}(x)$ de grado $n-1$ que sea la aproximación uniforme óptima de la función x^n en el intervalo $[-1, 1]$.

Resulta que el polinomio buscado verifica la ecuación

$$x^n - P_{n-1}(x) = \frac{1}{2^{n-1}} \cos n \arccos x. \quad (16)$$

Este hecho se deduce directamente del teorema de Chebichev, si probamos, primero, que el segundo miembro de (16) es un polinomio algebraico de grado n con el coeficiente de x^n igual a uno; segundo, que su valor absoluto sobre el intervalo $[-1, 1]$ alcanza su máximo, igual a $L = 1/2^{n-1}$, en los $n+1$ puntos $x_k = \cos k\pi/n$ ($k = 0, 1, \dots, n$), y tercero, que cambia de signo sucesivamente en estos puntos.

Que el segundo miembro de (16) es un polinomio de grado n con el coeficiente de x^n igual a uno se puede probar como sigue.

Supongamos que para un número natural determinado n hemos ya probado que

$$\begin{aligned} \cos n \arccos x &= 2^{n-1}[x^n - Q_{n-1}(x)]; \\ -\sqrt{1-x^2} \sin n \arccos x &= 2^{n-1}[x^{n+1} - Q_n(x)], \end{aligned}$$

donde Q_{n-1} y Q_n son polinomios algebraicos de grado $n-1$ y n , respectivamente. Entonces también serán válidas las ecuaciones análogas para $n+1$, como se establece fácilmente considerando las siguientes fórmulas:

$$\begin{aligned} \cos(n+1) \arccos x &= x \cos n \arccos x - \sqrt{1-x^2} \sin n \arccos x; \\ -\sqrt{1-x^2} \sin(n+1) \arccos x &= -x \sqrt{1-x^2} \sin n \arccos x + \\ &\quad + (x^2 - 1) \cos n \arccos x. \end{aligned}$$

Pero nuestras ecuaciones para $n = 1$ son ciertas, ya que

$$\begin{aligned}\cos \arccos x &= x, \\ -\sqrt{1-x^2} \sin \arccos x &= x^2 - 1.\end{aligned}$$

Por consiguiente, son ciertas para todo n .

El segundo miembro de (16) se denomina *polinomio Chebichev de grado n cuya desviación de cero es mínima*, ya que fue Chebichev el primero en anunciar y resolver este problema. Los primeros de estos polinomios son

$$\begin{aligned}T_0(x) &= 1, \\ T_1(x) &= x, \\ T_2(x) &= \frac{1}{2}(2x^2 - 1), \\ T_3(x) &= \frac{1}{4}(4x^3 - 3x), \\ T_4(x) &= \frac{1}{8}(8x^4 - 8x^2 + 1), \\ T_5(x) &= \frac{1}{16}(16x^5 - 20x^3 + 5x).\end{aligned}$$

Ya hemos visto el importante papel de los polinomios de Chebichev en cuestiones de interpolación y métodos de integración aproximada. Hagamos algunas observaciones acerca de la interpolación.

Del hecho de que la diferencia $f(x) - P_n(x)$ entre una función arbitraria $f(x)$ y su aproximación polinomial óptima cambie de signo en $n + 2$ puntos, se deduce (teniendo en cuenta las propiedades de las funciones continuas) que $P_n(x)$ coincide con $f(x)$ en $n + 1$ puntos específicos del intervalo $[a, b]$; es decir, $P_n(x)$ es un polinomio interpolador de grado n para $f(x)$ con una determinada elección de puntos de interpolación.

De este modo, el problema de la aproximación uniforme óptima de una función continua $f(x)$ se reduce a la elección, sobre el intervalo $[-1, 1]$, de un sistema $x_0, x_1, \dots, x_n$ de puntos de interpolación tal que el correspondiente polinomio interpolador de grado n tenga una desviación $\|f - Q\| = \max_x |f(x) - Q(x)|$ lo más pequeña posible. Desgraciadamente, los puntos de división requeridos son a

menudo difíciles de hallar en la práctica. Normalmente es necesario resolver el problema de una forma aproximada, y aquí es donde juegan un papel especial los polinomios de Chebichev. Resulta que si, en particular, se toman los ceros del polinomio $\cos(n+1) \arccos x$ (los puntos donde este polinomio es igual a cero) como puntos de interpolación, entonces el polinomio interpolador correspondiente —al menos para n suficientemente grande— dará una desviación uniforme de la función (si es ésta suficientemente regular) que difiere poco de la correspondiente desviación de la aproximación uniforme óptima de la función mediante un polinomio. La expresión algo vaga «difiere poco» se puede reemplazar, en muchos casos de especial importancia, por una estimación cuantitativa muy exacta, que no estableceremos aquí.

Volviendo al polinomio de Chebichev, considerémoslo en la forma $T_n(x) = M \cos n \arccos x$ ($-1 \leq x \leq 1$), donde M es un número positivo. Evidentemente, sobre el intervalo $[-1, 1]$ su valor absoluto no supera a M . Su derivada es

$$T'_n(x) = - \frac{nM \operatorname{sen} n \arccos x}{\sqrt{1-x^2}},$$

que, en el intervalo $[-1, 1]$, verifica la desigualdad

$$|T'_n(x)| \leq \frac{nM}{\sqrt{1-x^2}}.$$

Resulta que esta desigualdad es cierta para todo polinomio $P_n(x)$ de grado n que no supere a M en valor absoluto en el intervalo $[-1, 1]$; es decir, para la derivada de un polinomio de este tipo en el intervalo $[-1, 1]$ tenemos la desigualdad

$$|P'_n(x)| \leq \frac{nM}{\sqrt{1-x^2}}.$$

Esta desigualdad debe atribuirse a A. A. Markov, ya que se deduce directamente de resultados suyos que aún van algo más lejos. Markov la obtuvo en conexión con un problema que le sugirió D. I. Mendeleiev.

En 1912, S. N. Bernstein obtuvo una desigualdad similar, que lleva su nombre, para polinomios trigonométricos; y por medio de

estas desigualdades descubrió cómo establecer las propiedades de diferenciabilidad de una función si conocemos la rapidez con que converge a ella su sucesión de aproximaciones óptimas. Los resultados de este tipo, concernientes a funciones diferenciables, se verán en §6 y §7.

§6. *El teorema de Weierstrass; la aproximación óptima de una función referida a sus propiedades de diferenciabilidad*

El teorema de Weierstrass

Si aplicamos la definición general, dada en §4, de aproximación óptima de una función al caso de polinomios aproximadores, vamos a parar a la siguiente definición. La aproximación uniforme óptima de la función $f(x)$ en el intervalo $[a, b]$ mediante polinomios de grado n ocurre cuando el número (no negativo) $E_n(f)$ es igual al mínimo de la expresión

$$\max_{a \leq x \leq b} |f(x) - P_n(x)| = \|f - P_n\|,$$

tomada sobre todos los polinomios posibles $P_n(x)$ de grado n .

Independientemente de si podemos o no hallar el polinomio exacto que mejor aproxime la función dada $f(x)$, es de gran importancia tanto teórica como práctica estimar la cantidad $E_n(f)$ con la mayor exactitud posible. En realidad, si queremos aproximar la función f por un polinomio con error δ , en otras palabras, de tal forma que

$$|f(x) - P_n(x)| \leq \delta \quad (17)$$

para todo x del intervalo dado, entonces no tiene sentido elegirlo entre los polinomios de grado n para los que $E_n(f) > \delta$, ya que para este n indudablemente no existe ningún polinomio P_n que verifique (17). Por otra parte, si sabemos que $E_n(f) < \delta$, entonces es natural buscar para este n un polinomio $P_n(x)$ que aproxime a $f(x)$ con error δ , ya que tales polinomios existen evidentemente.

Las propiedades de la aproximación óptima de funciones de distintas clases han sido objeto de profundos y cuidadosos estudios. Ante todo observemos el importante hecho siguiente:

Si una función $f(x)$ es continua en el intervalo $[a, b]$, entonces su aproximación óptima $E_n(f)$ tiende a cero cuando n tiende a infinito.

Este es el teorema demostrado por Weierstrass a finales del pasado siglo. Tiene gran importancia, ya que garantiza la posibilidad de aproximar una función continua arbitraria por un polinomio con el error deseado. Como resultado de ello, el conjunto de todos los polinomios de cualquier grado guarda, respecto al conjunto de las funciones continuas sobre el intervalo, exactamente la misma relación que el conjunto Q de los números racionales respecto al conjunto R de los números reales (rationales e irracionales). En efecto, para cada número irracional α y cada número positivo arbitrariamente pequeño ϵ , se puede encontrar siempre un número racional r que verifica la desigualdad $|\alpha - r| < \epsilon$. Por otra parte, si $f(x)$ es una función continua sobre $[a, b]$ y ϵ es un número positivo arbitrariamente pequeño, entonces, por el teorema de Weierstrass, existe un polinomio algebraico $P_n(x)$ tal que para todo x del intervalo $[a, b]$ tenemos $|f(x) - P_n(x)| < \epsilon$. Por consiguiente, la aproximación óptima $E_n(f)$ de una función continua tiende a cero para $n \rightarrow \infty$.

Aclaremos el teorema de Weierstrass del siguiente modo. Dada la gráfica de una función continua arbitraria (Fig. 9) definida en el intervalo $[a, b]$, y un número positivo arbitrariamente pequeño ϵ , rodeemos nuestra gráfica con una banda de altura 2ϵ de tal forma que aquélla pase por el centro de la banda. Entonces siempre es posible elegir un polinomio algebraico

$$P_n(x) = a_0 + a_1x + \cdots + a_nx^n,$$

de grado suficientemente alto tal que su gráfica caiga por entero dentro de la banda.

Observemos lo siguiente. Como antes, sea $f(x)$ una función continua arbitraria sobre $[a, b]$, y sean $P_n(x)$ ($n = 1, 2, \dots$) los polinomios que constituyen la aproximación uniforme óptima. Es fácil ver que la función $f(x)$ se puede representar en forma de una serie $f(x) = P_1(x) + [P_2(x) - P_1(x)] + [P_3(x) - P_2(x)] + \cdots$, uniformemente convergente a $f(x)$ en $[a, b]$. Esto se deduce del hecho de que la suma de los n primeros términos de la serie es igual a $P_n(x)$, y

$$\max_{a \leq x \leq b} |f(x) - P_n(x)| = E_n(f),$$

mientras $E_n(f) \rightarrow 0$ cuando $n \rightarrow \infty$.

Como resultado tenemos una nueva formulación del teorema de Weierstrass:

Toda función continua en el intervalo $[a, b]$ se puede representar por una serie de polinomios algebraicos que converge uniformemente a la función.

Este resultado tiene gran importancia teórica. Garantiza la posibilidad de representar una función continua arbitraria, cualquiera que sea la forma en que venga dada (por ejemplo, mediante una gráfica), en forma de una expresión analítica. (Por expresión analítica entendemos una función elemental o una función derivada de una sucesión de funciones elementales mediante un proceso de paso al límite.) Históricamente este resultado destruyó finalmente la noción de expresión analítica que había existido en matemáticas casi hasta mediados del siglo pasado. Decimos «finalmente» porque el teorema de Weierstrass estuvo precedido por una serie de resultados generales de tipo similar, relacionados sobre todo con las series de Fourier. Antes de que se obtuvieran estos resultados se suponía que las expresiones analíticas constituían el medio de representar las propiedades especialmente deseables que eran características de las funciones analíticas. Por ejemplo, se solía dar por supuesto que las expresiones analíticas eran infinitamente diferenciables y podían incluso ser desarrolladas en serie de potencias. Pero estas ideas demostraron no tener fundamento. Una función puede no tener derivada en su intervalo de definición y sin embargo ser representable por medio de una expresión analítica.

Desde un punto de vista metodológico, el valor de este descubrimiento radica en el hecho de que nos permite comprender con absoluta claridad que, al menos en principio, los métodos de las matemáticas son aplicables a una clase mucho más amplia de leyes de lo que antes se creía.

En la actualidad se conocen muchas demostraciones distintas del teorema de Weierstrass. La mayor parte de ellas se reducen a la construcción de una sucesión de polinomios para una función continua dada f que converja a ella uniformemente. El polinomio

$$B_n(x) = \sum_{k=1}^n C_n^k x^k (1-x)^{n-k} f\left(\frac{k}{n}\right),$$

aproximará una función continua $f(x)$ sobre el intervalo $[0,1]$. Se denomina *polinomio de Bernstein*. Al crecer n , este polinomio converge uniformemente en el intervalo $[0,1]$ hacia la función que lo

ha engendrado ⁶. Aquí C_n^k es el número de combinaciones de n elementos tomados de k en k .

Observemos que un teorema análogo al de Weierstrass es válido en el campo complejo. Resultados exhaustivos en este sentido se deben a M. A. Laurentiev, M. V. Keldysch, y S. N. Mergelian.

Relación entre el orden de la aproximación uniforme óptima de una función y sus propiedades de diferenciabilidad

Señalemos aún los siguientes resultados. Si una función $f(x)$ tiene en el intervalo $[a, b]$ una derivada $f^{(r)}(x)$ de orden r que no supera al número K en valor absoluto, entonces su aproximación óptima $E_n(f)$ verifica la desigualdad

$$E_n(f) \leq \frac{c_r K}{n^r}, \quad (18)$$

donde c_r es una constante que depende solamente de r (teorema de Jackson). Por la desigualdad (18) se puede ver que cuando crece n , la cantidad $E_n(f)$ converge a cero más rápidamente para funciones con derivadas de orden más alto. Dicho con otras palabras, cuanto mejor (más regular) sea la función, más rápidamente converge a cero su aproximación óptima. Bernstein demostró que en cierto sentido la inversa de esta proposición es también cierta.

Mejores aún en este aspecto que las funciones diferenciables son las funciones analíticas. Bernstein demostró que para tales funciones $E_n(f)$ verifica la desigualdad

$$E_n(f) \leq cq^n, \quad (19)$$

donde c y q son constantes que sólo dependen de la función f , y $0 < q < 1$; es decir, $E_n(f)$ converge a cero más rápidamente que una cierta serie geométrica decreciente. También demostró que, recíprocamente, la desigualdad (19) implica que la función f es analítica en $[a, b]$.

Hemos dado algunos resultados muy importantes, que fueron descubiertos a principios de este siglo y que han sido característicos de la dirección tomada por las modernas investigaciones en la teoría de aproximación de funciones. El valor práctico de estos resultados se puede ver en el siguiente ejemplo.

⁶ Debemos observar que, a pesar de su simplicidad, los polinomios de Bernstein son poco utilizados en la práctica. La explicación es que convergen muy despacio, incluso para funciones con buenas propiedades de diferenciabilidad.

Si $Q_n(x)$ es un polinomio de grado n , que interpola la función $f(x)$ en el intervalo $[-1, 1]$ en los $n + 1$ puntos de interpolación que corresponden a los ceros del polinomio de Chebichev $\cos(n + 1) \arccos x$, entonces, sobre este intervalo, se tiene la desigualdad $|f(x) - Q_n(x)| < c \ln n E_n(f)$, donde c es una constante independiente de n , y $E_n(f)$ la aproximación óptima de f en el intervalo $[-1, 1]$. En esta desigualdad podemos reemplazar $E_n(f)$ por las expresiones mayores que aparecen en (18) o (19), supuesto que f sea suficientemente regular, obteniendo una buena estimación de la aproximación de nuestro polinomio interpolador. Como $\ln n$ crece muy lentamente al aumentar n , el orden de la estimación en este caso difiere poco del orden de la convergencia a cero de $E_n(f)$. La ventaja de la interpolación por los puntos de Chebichev consiste en que para otros puntos de interpolación el factor $c \ln n$, en la correspondiente desigualdad, está reemplazado por un factor que crece más rápidamente; esto es particularmente exacto en el caso de puntos de interpolación equidistantes.

§7. Series de Fourier

El origen de las series de Fourier

Las series de Fourier surgieron en relación con el estudio de ciertos fenómenos físicos, concretamente pequeñas oscilaciones de medios elásticos. Un ejemplo característico es la oscilación de una cuerda musical. En realidad, la investigación de cuerdas vibrantes fue el origen histórico de las series de Fourier y determinó la dirección en que se desarrolló esta teoría.

Consideremos (Fig. 11) una cuerda tensa, con extremos fijos en

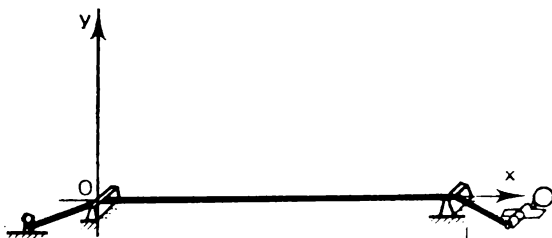


FIG. 11

los puntos $x = 0$ y $x = l$ del eje Ox . Si desplazamos la cuerda de su posición de equilibrio, oscilará.

Sigamos el movimiento de un punto específico de la cuerda de abscisa x_0 . Su desviación vertical del punto de equilibrio es una función $\phi(t)$ del tiempo. Se puede demostrar que siempre es posible dar a la cuerda una posición y velocidad iniciales en $t = 0$ tales que el punto que hemos convenido seguir describa oscilaciones armónicas en la dirección vertical, definida por la función

$$\phi = \phi(t) = A \cos akt + B \operatorname{sen} akt. \quad (20)$$

Aquí a es una constante que depende sólo de las propiedades físicas de la cuerda (densidad, tensión y longitud), k es un número arbitrario, y A y B son constantes.

Observemos que nuestra discusión se refiere solamente a pequeñas oscilaciones de la cuerda. Esto nos autoriza a suponer que cada punto x_0 oscila solamente en la dirección vertical, ignorando los desplazamientos según la horizontal⁷. Suponemos también que la fricción originada por la oscilación de la cuerda es tan pequeña que podemos despreciarla. Como consecuencia de estas hipótesis aproximadas, las vibraciones no se extinguirán.

Las posibilidades de vibración del punto x_0 no se agotan, naturalmente, en los movimientos periódicos definidos por las funciones armónicas (20); estas funciones tienen la notable propiedad siguiente. Los experimentos y la teoría correspondiente prueban que cada oscilación posible del punto x_0 es el resultado de la combinación de ciertas oscilaciones armónicas de la forma (20). Se obtienen oscilaciones relativamente sencillas combinando un número finito de aquéllas; es decir, vienen descritas por funciones de la forma

$$\phi(t) = A_0 + \sum_{k=1}^n (A_k \cos akt + B_k \operatorname{sen} akt),$$

donde A_k y B_k son constantes. Estas funciones se denominan polinomios trigonométricos. En casos más complicados, la oscilación será

⁷ Esta cuestión está directamente relacionada con la ecuación diferencial de la cuerda vibrante

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \left(a = \frac{l}{\pi} a \right),$$

que se discutió en el capítulo 6.

el resultado de combinar un número infinito de vibraciones de la forma (20), correspondientes a $k = 1, 2, 3, \dots$ y con constantes convenientemente elegidas A_k y B_k que dependen del número k . Por consiguiente, llegamos a la necesidad de representar una función dada $\phi(t)$ de período $2\pi/\alpha$, que describe una oscilación arbitraria del punto x_0 , mediante una serie

$$\phi(t) = A_0 + \sum_{k=1}^{\infty} (A_k \cos akt + B_k \operatorname{sen} akt). \quad (21)$$

Existen otras muchas situaciones físicas en las que es natural considerar una función dada —aun cuando no describa necesariamente una oscilación— como suma de una serie trigonométrica infinita de la forma (21). Tal caso se da, por ejemplo, en relación con la misma cuerda vibrante. La ley exacta para la oscilación subsiguiente de una cuerda, a la que hemos dado al principio un desplazamiento inicial específico (por ejemplo, como se ilustra en la Fig. 12), es fácil de calcular, siempre que conozcamos el desarrollo en serie trigonométrica $f(x) = \sum_{k=1}^{\infty} a_k \operatorname{sen} (k\pi/l)x$ —un caso particular de la serie (21)— de la función $f(x)$ que describe la posición inicial.

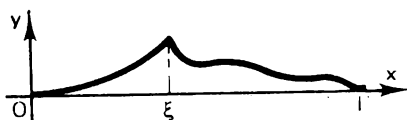


FIG. 12

Desarrollo de funciones en serie trigonométrica

Sobre la base de lo que ya se ha dicho surge la siguiente cuestión fundamental: ¿qué funciones de período $2\pi/\alpha$ pueden representarse como suma de una serie trigonométrica de la forma (21)? Esta cuestión fue planteada en el siglo XVIII por Euler y Bernoulli en relación con el estudio de este último de la cuerda vibrante. Bernoulli adoptó aquí el punto de vista, sugerido por consideraciones físicas, de que una clase muy amplia de funciones, incluidas todas las gráficas trazadas a mano, se pueden desarrollar en serie trigonométrica. Esta opinión fue acogida duramente por muchos contemporáneos de Bernoulli, aferrándose tenazmente a la idea de que si una función se representa como una expresión analítica (tal como una serie trigonométrica), entonces debe tener buenas propiedades de diferenciabilidad. Pero la función ilustrada en la figura 12 ni siquiera tiene deri-

vada en el punto ξ ; en tal caso, ¿cómo se puede caracterizar mediante una única expresión analítica en todo el intervalo $[0, l]$?

Ahora sabemos que el punto de vista físico de Bernoulli era bastante correcto. Pero para poner fin a la controversia fue necesario esperar un siglo entero, ya que la respuesta definitiva a estas cuestiones requería ante todo que los conceptos de límite y de suma de una serie se formularan sobre una base rigurosa.

Las investigaciones matemáticas fundamentales que confirmaron

el punto de vista físico, pero que estaban basadas en ideas más antiguas, relativas a los fundamentos del análisis, fueron completadas en 1807-1822 por el matemático francés Fourier.

Finalmente, en 1829, el matemático alemán Dirichlet demostró, con

el mismo rigor con que se procede en las matemáticas actuales, que toda función continua de período $2\pi/a$ ⁸ que en cualquier período tenga un número finito de máximos y mínimos se puede desarrollar en una única serie trigonométrica de Fourier uniformemente convergente⁹ a la función.

La figura 13 ilustra una función que verifica las condiciones de Dirichlet. Su gráfica es continua y periódica, con período 2π , y tiene un máximo y un mínimo en el período $0 \leq x \leq 2\pi$.

Coefficientes de Fourier

En lo que sigue consideraremos funciones de período 2π , con lo cual se simplifican las fórmulas. Sea $f(x)$ una función continua de período 2π que verifique la condición de Dirichlet. Por el teorema del mismo nombre podemos desarrollarla en una serie trigonométrica

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \operatorname{sen} kx), \quad (22)$$

⁸ La función $f(x)$ tiene período ω si verifica la ecuación $f(x + \omega) = f(x)$.

⁹ En realidad, el teorema de Dirichlet también se aplica a cierta clase de funciones discontinuas, las llamadas funciones de variación acotada. Para funciones discontinuas, por supuesto, la serie correspondiente no es uniformemente convergente.

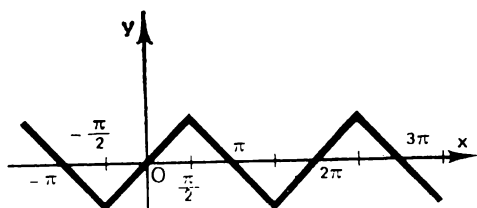


FIG. 13

que converge uniformemente a ella. El hecho de que el primer término se escriba como $a_0/2$ en vez de a_0 no tiene otro significado que el de la pura conveniencia, como veremos más tarde.

Planteemos el problema siguiente: calcular los coeficientes a_k y b_k de la serie para una función dada $f(x)$.

Para ello consideramos las siguientes ecuaciones:

$$\begin{aligned}\int_{-\pi}^{\pi} \cos kx \cos lx \, dx &= 0 \quad (k \neq l; k, l = 0, 1, \dots), \\ \int_{-\pi}^{\pi} \sin kx \sin lx \, dx &= 0 \quad (k \neq l; k, l = 0, 1, \dots), \\ \int_{-\pi}^{\pi} \sin kx \cos lx \, dx &= 0 \quad (k, l = 0, 1, 2, \dots), \\ \int_{-\pi}^{\pi} \cos^2 kx \, dx &= \pi \quad (k = 1, 2, \dots), \\ \int_{-\pi}^{\pi} \sin^2 kx \, dx &= \pi \quad (k = 1, 2, \dots),\end{aligned}\tag{23}$$

que el lector puede comprobar. Estas integrales son fáciles de calcular reduciendo los productos de las distintas funciones trigonométricas a sumas y diferencias, y sus cuadrados, a expresiones que contengan las correspondientes funciones trigonométricas del ángulo doble. La primera ecuación establece que la integral, extendida a un período de la función, del producto de dos funciones distintas de la sucesión $1, \cos x, \sin x, \cos 2x, \sin 2x, \dots$, es igual a cero (la llamada propiedad de ortogonalidad de las funciones trigonométricas). Por otra parte, la integral del cuadrado de cada una de las funciones de esta sucesión es igual a π . La primera función, idénticamente igual a uno, es una excepción, ya que la integral de su cuadrado extendida al período es igual a 2π . Por esta razón, conviene escribir el primer término de la serie (22) en la forma $a_0/2$.

Ahora podemos resolver fácilmente el problema. Para calcular el coeficiente a_m , multiplicamos el primer miembro y cada término del segundo de la serie (22) por $\cos mx$ e integramos término a término a lo largo de un período 2π , lo cual está permitido puesto que la serie obtenida después de multiplicar por $\cos mx$ es uniformemente

convergente. Por (23), todas las integrales del segundo miembro, con excepción de la correspondiente a $\cos mx$, serán nulas, por lo que evidentemente

$$\int_{-\pi}^{\pi} f(x) \cos mx \, dx = a_m \pi,$$

en donde

$$a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos mx \, dx \quad (m = 0, 1, 2, \dots). \quad (24)$$

Análogamente, multiplicando ambos miembros de (22) por $\sin mx$ e integrando a lo largo del período, obtenemos una expresión para los coeficientes

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin mx \, dx \quad (m = 1, 2, \dots), \quad (25)$$

con lo que hemos resuelto el problema. Los números a_m y b_m , calculados mediante las fórmulas (24) y (25), se denominan *coeficientes de Fourier* de la función $f(x)$.

Tomemos como ejemplo la función $f(x)$ de período 2π ilustrada en la figura 13. Evidentemente esta función es continua y verifica la condición de Dirichlet, por lo que su serie de Fourier converge a ella uniformemente.

Es fácil ver que esta función también verifica la condición $f(-x) = -f(x)$. Está claro que esta misma condición también vale para la función $F_1(x) = f(x) \cos mx$, lo que significa que la gráfica de $F_1(x)$ es simétrica respecto al origen. Por razonamientos geométricos se ve que $\int_{-\pi}^{\pi} F_1(x) \, dx = 0$, por lo que $a_m = 0$ ($m = 0, 1, 2, \dots$). Además, no es difícil ver que las funciones $F_2(x) = f(x) \sin mx$ tienen una gráfica simétrica con respecto al eje Oy , por lo que

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} F_2(x) \, dx = \frac{2}{\pi} \int_0^{\pi} F_2(x) \, dx.$$

Pero para m par esta gráfica es simétrica con respecto al centro $\pi/2$ del segmento $[0, \pi]$, de modo que $b_m = 0$ para m par. Si m es impar,

$m = 2l + 1$ ($l = 0, 1, 2, \dots$), la gráfica de $F_2(x)$ es simétrica con respecto a la recta $x = \pi/2$, por lo que

$$b_{2l+1} = \frac{4}{\pi} \int_0^{\pi/2} F_2(x) dx.$$

Ahora bien, como se puede ver en el esquema, sobre el segmento $[0, \pi/2]$ tenemos simplemente $f(x) = x$, por lo que, integrando por partes, obtenemos

$$b_{2l+1} = \frac{4}{\pi} \int_0^{\pi/2} x \operatorname{sen} (2l + 1)x dx = \frac{4(-1)^l}{\pi(2l + 1)^2},$$

y por consiguiente

$$f(x) = \frac{4}{\pi} \sum_{l=1}^{\infty} \frac{(-1)^l \operatorname{sen} (2l + 1)x}{(2l + 1)^2}$$

De este modo hemos hallado el desarrollo de nuestra función en serie de Fourier.

Convergencia de las sumas parciales de Fourier a la función generatriz

En las aplicaciones se suele tomar como aproximación de la función de período 2π la suma

$$S_n = \frac{a_0}{2} + \sum_1^n (a_k \cos kx + b_k \operatorname{sen} kx)$$

de los n primeros términos de su serie de Fourier, planteándose entonces la cuestión del error de la aproximación. Si la función $f(x)$ de período 2π tiene una derivada $f^{(r)}(x)$ de orden r que verifique, para todo x , la desigualdad $|f^{(r)}(x)| \leq K$, entonces el error de la aproximación se puede estimar como sigue:

$$|f(x) - S_n(x)| \leq \frac{c_r K \ln n}{n^r},$$

donde c_r es una constante que sólo depende de r . Vemos que el error tiende a cero cuando n crece, y la convergencia es tanto más rápida cuantas más derivadas tenga la función.

Para una función analítica sobre toda la recta real existe una estimación mejor aún:

$$|f(x) - S_n(x)| < cq^n, \quad (26)$$

donde c y q son constantes positivas que dependen solamente de f y $q < 1$. El recíproco también es cierto, es decir, que, si la desigualdad (26) se verifica para una función dada, entonces es necesariamente analítica. Este hecho, que fue descubierto a principios del presente siglo, pone fin en cierto sentido a la controversia entre D. Bernoulli y sus contemporáneos. Ahora podemos enunciar: el hecho de que una función sea desarrollable en una serie de Fourier que converja a ella no implica que la función sea analítica; sin embargo, será analítica si su desviación respecto de la suma de los n primeros términos de la serie de Fourier decrece más rápidamente que los términos de alguna serie geométrica decreciente.

La comparación de las estimaciones de las aproximaciones proporcionadas por las sumas de Fourier, por un lado, y las correspondientes estimaciones para las aproximaciones óptimas de las mismas funciones por polinomios trigonométricos, por otro, prueba que para funciones regulares las sumas de Fourier dan aproximaciones muy buenas, que están, de hecho, próximas a las aproximaciones óptimas. Pero para funciones continuas no regulares la situación es peor: entre ellas hay algunas funciones cuya serie de Fourier diverge en el conjunto de todos los puntos racionales.

Señalemos que en la teoría de series de Fourier hay una cuestión que se planteó hace mucho tiempo y que aún está sin respuesta: ¿existe una función periódica continua cuya serie de Fourier no converja para ningún x a la función cuando $n \rightarrow \infty$? El mejor resultado en este sentido se debe a A. N. Kolmogorov, quien demostró en 1926 que existe una función periódica integrable Lebesgue cuya serie de Fourier no converge a ella en ningún punto. Pero una función integrable Lebesgue puede no ser continua, como es el caso de la función construida por Kolmogorov. El problema aún espera su solución final.

Para conseguir aproximaciones de funciones periódicas continuas arbitrarias mediante polinomios trigonométricos se utilizan actualmente los llamados métodos de sumación de series de Fourier. En lugar de las sumas de Fourier como aproximación de una función dada, consideramos ciertas modificaciones de ellas. Un método muy simple de esta clase fue propuesto por el matemático húngaro Fejér.

Para una función periódica continua construimos en primer lugar, de una forma puramente formal, su serie de Fourier, que puede ser divergente, y a continuación formamos la media aritmética de las n primeras sumas parciales

$$\sigma_n(x) = \frac{S_0(x) + S_1(x) + \cdots + S_n(x)}{n+1}. \quad (27)$$

Esta es la llamada *suma de Fejér* de orden n correspondiente a la función dada $f(x)$. Fejér demostró que cuando $n \rightarrow \infty$ esta suma converge uniformemente a $f(x)$.

§8. Aproximación en el sentido del cuadrado medio

Volvamos al problema de la cuerda vibrante. Supongamos que en un cierto instante t_0 la cuerda tiene la forma $y = f(x)$. Podemos probar que su energía potencial W —el trabajo que puede realizar cuando se mueve desde la posición dada a la de equilibrio— es igual (para desviaciones pequeñas de la cuerda) a la integral $W = \int_0^l f'^2(x) dx$, salvo un factor constante. Supongamos ahora que queremos aproximar la función $f(x)$ por otra $\phi(x)$. Junto con la cuerda dada, consideraremos otra cuya forma esté caracterizada por $\phi(x)$, y una tercera aún, definida por la función $f(x) - \phi(x)$. Se puede demostrar que, si la energía

$$\int_0^l [f'(x) - \phi'(x)]^2 dx \quad (28)$$

de la tercera cuerda es pequeña, entonces la diferencia entre las energías de las dos primeras es también pequeña¹⁰. Si nos interesa que la segunda cuerda tenga una energía que difiera poco de la primera,

¹⁰ En efecto, si

$$\int_0^l f'^2 dx \leq M^2 \text{ y } \int_0^l \phi'^2 dx \leq M^2,$$

entonces

$$\begin{aligned} & \left| \int_0^l f'^2 dx - \int_0^l \phi'^2 dx \right| \leq \\ & \leq \left(\sqrt{\int_0^l f'^2 dx} + \sqrt{\int_0^l \phi'^2 dx} \right) \left| \sqrt{\int_0^l f'^2 dx} - \sqrt{\int_0^l \phi'^2 dx} \right| \leq 2M \sqrt{\int_0^l (f' - \phi')^2 dx}. \end{aligned}$$

debemos intentar hallar una función $\phi'(x)$ para la cual la integral (28) sea lo más pequeña posible. Llegamos de este modo al problema de aproximación de una función —en este caso $f'(x)$ — en el sentido del cuadrado medio.

Veamos cómo se plantea este problema en el caso general. Sea $F(x)$ una función definida en el intervalo $[a, b]$ y otra función

$$\Phi(x; \alpha_0, \alpha_1, \dots, \alpha_n), \quad (29)$$

que depende, no sólo de x , sino también de los parámetros $\alpha_0, \alpha_1, \dots, \alpha_n$. Es necesario elegir estos parámetros de forma que minimicen la integral

$$\int_a^b [F(x) - \Phi(x; \alpha_0, \alpha_1, \dots, \alpha_n)]^2 dx. \quad (30)$$

La idea de este problema es muy similar a la del problema de Chebichev. Aquí también se trata de hallar la aproximación óptima de la función $F(x)$ mediante funciones de la familia (29), pero solamente en el sentido del cuadrado medio. Ahora no nos importa que la diferencia $F - \Phi$ sea pequeña o grande para todos los puntos x del intervalo $[a, b]$; sobre una parte pequeña del intervalo la diferencia $F - \Phi$ puede incluso ser grande con tal que la integral (30) sea

pequeña, como en el caso, por ejemplo, de las dos gráficas ilustradas en la figura 14. La pequeñez de la cantidad (30) prueba que las funciones F y Φ están próximas una a otra en la mayor parte del intervalo¹¹. En cuanto a la elección en la práctica de un método u

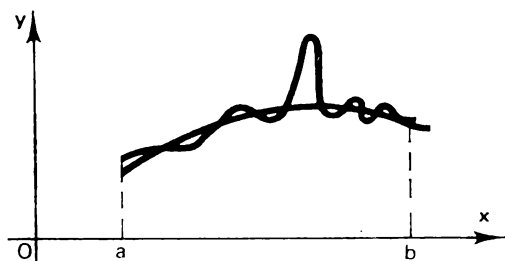


FIG. 14

otro de aproximación, todo depende de nuestro propósito. En el ejemplo de la cuerda, es natural aproximar la función $f'(x)$ en el sentido del cuadrado medio. Por otra parte, el método del cuadrado medio era

¹¹ En el capítulo 19 veremos que existe una profunda analogía entre la proximidad de las funciones en el sentido del cuadrado medio y la distancia entre puntos en el espacio ordinario.

poco satisfactorio para Chebichev al tratar de resolver problemas de la construcción de mecanismos, ya que una máquina con una pieza diseñada más allá de los límites de tolerancia sería inadmisibile: tal diseño estropearía toda la máquina. De este modo, Chebichev tuvo que desarrollar un nuevo método matemático correspondiente al problema con que se enfrentó. Podríamos decir que, desde el punto de vista del cálculo, el método del cuadrado medio es más conveniente, ya que se puede reducir a la aplicación de métodos bien desarrollados del análisis general.

Consideremos, por ejemplo, el siguiente problema característico.

Queremos obtener la aproximación óptima, en el sentido del cuadrado medio, de una función dada $f(x)$ sobre el intervalo $[a, b]$ mediante sumas de la forma

$$\sum_1^n a_k \phi_k(x),$$

donde los a_k son constantes y las funciones $\phi_k(x)$ son continuas y forman un sistema ortonormal.

Esto último significa que se verifican las siguientes ecuaciones:

$$\int_a^b \phi_k \phi_l dx = 0 \quad k \neq l \quad (k, l = 1, 2, \dots, n),$$

$$\int_a^b \phi_k^2 dx = 1 \quad (k = 1, 2, \dots, n).$$

Introduzcamos los números

$$a_k = \int_a^b f(x) \phi_k(x) dx \quad (k = 1, \dots, n).$$

Estos números a_k se denominan coeficientes de Fourier de f respecto de las ϕ_k .

Para coeficientes arbitrarios a_k , en virtud de las propiedades de ortonormalidad de ϕ_k tenemos la ecuación

$$\begin{aligned} \int_a^b \left(f - \sum_1^n a_k \phi_k \right)^2 dx &= \int_a^b f^2 dx + \sum_1^n a_k^2 - 2 \sum_1^n a_k a_k = \\ &= \left(\int_a^b f^2 dx - \sum_1^n a_k^2 \right) + \sum_1^n (a_k - a_k)^2. \end{aligned}$$

El primer término del segundo miembro de la ecuación obtenida no depende de los a_k . De este modo el segundo miembro será mínimo para aquellos a_k que hagan mínimo el segundo sumando, y evidentemente esto sólo puede ocurrir si los a_k coinciden con los correspondientes coeficientes de Fourier a_k .

Así, hemos llegado al siguiente resultado: si las funciones ϕ_k forman un sistema ortonormal sobre el intervalo $[a, b]$, entonces la suma $\sum_1^n a_k \phi_k(x)$ será la aproximación óptima, en el sentido del cuadrado medio, de la función $f(x)$ en este intervalo si, y sólo si, los números a_k son los coeficientes de Fourier de la función f con respecto a $\phi_k(x)$.

En virtud de la ecuación (23) se establece fácilmente que las funciones

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots$$

forman un sistema ortonormal sobre el intervalo $[0, 2\pi]$. Así, la proposición enunciada, aplicada a las funciones trigonométricas, tendrá la siguiente forma.

La suma de Fourier $S_n(x)$, calculada para una función continua dada $f(x)$ de período 2π , es la aproximación óptima, en el sentido del cuadrado medio, de la función $f(x)$ en el intervalo $[0, 2\pi]$, entre todos los polinomios trigonométricos

$$t_n(x) = a_0 + \sum_1^n (a_k \cos kx + \beta_k \sin kx)$$

de orden n .

A partir de este resultado y del teorema de Fejér, formulado en el §7, llegamos a otro hecho notable.

Sea $f(x)$ una función continua de período 2π , y $\sigma_n(x)$, su suma de Féjer de orden n , definida en el §7 por la ecuación (27).

Introducimos la notación

$$\max |f(x) - \sigma_n(x)| = \eta_n.$$

Como las sumas de Fourier $S_k(x)$ ($k = 0, 1, \dots, n$) son polinomios trigonométricos de orden $k \leq n$, es evidente que $\sigma_n(x)$ es un polino-

mio trigonométrico de orden n . Así, de la propiedad minimal de la suma demostrada anteriormente, tenemos la desigualdad

$$\int_{-\pi}^{\pi} [f(x) - S_n(x)]^2 dx \leq \int_{-\pi}^{\pi} [f(x) - \sigma_n(x)]^2 dx \leq \int_{-\pi}^{\pi} \eta_n^2 dx = 2\pi \eta_n^2.$$

Como, por el teorema de Fejér, la cantidad η_n converge a cero para $n \rightarrow \infty$, obtenemos el importante resultado siguiente.

Para una función continua de período 2π tenemos la ecuación

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} [f(x) - S_n(x)]^2 dx = 0.$$

En este caso decimos que la suma de Fourier de orden n de una función continua $f(x)$ converge a $f(x)$ en el sentido del cuadrado medio cuando n crece indefinidamente.

En realidad, esta afirmación es cierta para una clase más amplia de funciones: aquellas que son integrables, junto con sus cuadrados, en el sentido de Lebesgue.

Nos detendremos aquí, omitiendo otros hechos interesantes de la teoría de series de Fourier y funciones ortogonales, basados en la aproximación en el sentido del cuadrado medio. En el capítulo 6 se han introducido ya importantes aplicaciones físicas de sistemas ortogonales de funciones. Finalmente, observemos que estas cuestiones se discutirán también desde un punto de vista algo diferente en el capítulo 19.

Capítulo 13

MÉTODOS DE APROXIMACIÓN Y TÉCNICAS DE CÁLCULO

§1. *Aproximación y métodos numéricos*

Peculiaridades distintivas de los métodos de aproximación

En muchos casos la aplicación de las matemáticas al estudio de acontecimientos del mundo exterior se basa en el hecho de que las leyes que rigen estos acontecimientos tienen un carácter cuantitativo y se pueden describir mediante ciertas fórmulas, ecuaciones o desigualdades. Esto nos permite investigar los acontecimientos numéricamente y hacer los cálculos que son tan necesarios en la vida práctica.

Tan pronto como se haya hallado una ley cuantitativa, se pueden utilizar métodos puramente matemáticos para investigarla. Para precisar, tomemos una ley descrita por una ecuación: puede ser la ley del movimiento de un cuerpo en mecánica newtoniana, la ley de conducción del calor o la de propagación de ondas electromagnéticas, etcétera. Tales ecuaciones se han discutido con detalle en los capítulos 5 y 6. Normalmente, la ecuación lleva asociadas ciertas condiciones que debe verificar su solución (en los capítulos 5 y 6 son las condiciones iniciales y de contorno) y que determinan una solución única.

La primera y más importante tarea matemática será entonces la siguiente:

1. Establecer la existencia de una solución. Aunque parezca evidente desde el punto de vista físico que el problema tiene una solución, la demostración matemática de la resolubilidad se considera en general como el testimonio necesario de que la formulación del problema es satisfactoria. En una amplia clase de problemas es posible establecer matemáticamente la existencia de una solución.

2. Intentar hallar una expresión explícita o *fórmula* para la cantidad que caracteriza el acontecimiento en cuestión. Normalmente tal expresión solamente puede hallarse en los casos más simples. A menudo sucede que la expresión explícita obtenida es tan complicada que resulta muy difícil e incluso imposible utilizarla para obtener los resultados numéricos deseados.

3. Encontrar un procedimiento para construir una *fórmula de aproximación* que dé una solución con el grado de exactitud deseado. Esto puede hacerse en muchos casos.

4. Pero muy a menudo será posible hallar uno o más métodos de *cálculo numérico* directo de la solución.

El desarrollo de tales métodos numéricos (muchos de los cuales son aproximados) de resolver problemas científicos y tecnológicos ha dado lugar a una rama particular de las matemáticas que actualmente se ha dado en llamar matemáticas de computación.

Los métodos de esta rama son naturalmente aproximados, ya que cada cantidad se calcula solamente con un cierto número de cifras significativas; por ejemplo, con cinco, seis, etc., cifras decimales.

Para las aplicaciones esto es suficiente, puesto que conocer el valor exacto de una cantidad suele ser innecesario. En cuestiones técnicas, por ejemplo, la cantidad buscada sirve normalmente para determinar las dimensiones u otros parámetros del artículo manufacturado. Todo proceso de fabricación es siempre aproximado, por lo que los cálculos técnicos cuya exactitud sobrepase las «tolerancias» permitidas, evidentemente carecerán de valor.

Así, a efectos de cálculo no son necesarias fórmulas exactas ni soluciones exactas de las ecuaciones. Las fórmulas y soluciones exactas se pueden reemplazar por otras reconocidamente inexactas, con tal que éstas sean tan fieles a las originales que el error cometido no sobrepase ciertos límites.

Después volveremos sobre esta cuestión de sustituir un problema por otro. De momento queremos poner de relieve el primer rasgo característico de los métodos de cálculo: si bien por su misma naturaleza, no pueden, por regla general, producir más que resultados

aproximados, lo cierto es que en la práctica solamente se necesitan tales resultados.

Dirijamos ahora nuestra atención a un segundo aspecto de los métodos de cálculo en matemáticas. En cualquier cálculo sólo podemos operar con un número finito de dígitos y obtener los resultados después de un número finito de operaciones aritméticas. Si realizamos los cálculos de acuerdo con cierta fórmula, entonces esta última se debe transformar primero de tal modo que incluya sólo un número finito de términos con un número finito de parámetros. Es sabido, por ejemplo, que muchas funciones se pueden representar como la suma de una serie de potencias.

$$f(x) = c_0 + c_1x + c_2x^2 + \dots \quad (1)$$

Así, la función $\sin x$, donde x es la medida en radianes de un ángulo, se puede desarrollar en la serie de potencias

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

Para hallar el valor exacto de $f(x)$, sería necesario sumar «todos» los términos de la serie (1), pero, generalmente, esto es imposible. Para hallar $f(x)$ aproximadamente basta con tomar un cierto número finito de términos de la serie. Por ejemplo, se puede demostrar que para calcular $\sin x$ con un error menor que 10^{-5} para un ángulo comprendido entre cero y la mitad de un ángulo recto basta con tomar los términos hasta x^5 , sustituyendo $\sin x$ por el polinomio.

$$\frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

Para la resolución numérica de un problema de análisis matemático que consista en determinar cierta función, debemos reemplazarlo de algún modo por el problema de hallar ciertos parámetros numéricos, cuyo conocimiento nos permita efectuar un cálculo aproximado de la función incógnita. Aclararemos esto con un ejemplo.

Queremos resolver, en el intervalo $a \leq x \leq b$, el problema de contorno para la ecuación diferencial

$$L(y) - f(x) = y'' + p(x)y' + q(x)y - f(x) = 0 \quad (2)$$

con condiciones de contorno $y(a) = 0$, $y(b) = 0$. En uno de los posibles métodos de resolverlo, denominado método de Galerkin, partimos de un sistema de funciones linealmente independientes $\omega_1(x)$, $\omega_2(x)$, $\dots$, que verifican las condiciones de contorno (capítulo 6, §5). Este sistema se elige de modo que sea «completo» en el sentido de que una función que sea integrable en $[a, b]$ y ortogonal a todas las ω_k ($k = 1, 2, \dots$) será igual a cero en todos (más exactamente, en «casi todos») los puntos del intervalo. La condición de que $y(x)$ verifique la ecuación diferencial (2) se puede escribir en forma de una relación de ortogonalidad

$$\int_a^b [L(y) - f] \omega_k dx = 0 \quad (k = 1, 2, \dots). \quad (3)$$

Supongamos que la solución del problema se puede desarrollar en serie en las ω_k :

$$y(x) = a_1 \omega_1(x) + a_2 \omega_2(x) + \dots \quad (4)$$

Intentemos determinar las condiciones que deben verificar los coeficientes a_k . Para a_k arbitrarios la suma de la serie (4) verificará las condiciones de contorno. Falta elegir los a_k de manera que se verifiquen las ecuaciones (3). Los coeficientes a_k forman una familia infinita y calcularlos todos es generalmente imposible. Para simplificar, conservamos solamente un número finito de términos del segundo miembro de (4), obteniendo así la expresión

$$y(x) \approx a_1 \omega_1(x) + \dots + a_n \omega_n(x). \quad (5)$$

No podemos esperar satisfacer la ecuación (3) para todos los ω_k ($k = 1, 2, \dots$), ya que tenemos solamente n parámetros arbitrarios a_k ($k = 1, 2, \dots, n$). De este modo nos vemos obligados a renunciar a una solución exacta de la ecuación diferencial (2). Pero es natural esperar que la suma (5) verifique esta ecuación diferencial con un error pequeño si tomamos n suficientemente grande y si la condición (3) se verifica para las n primeras funciones ω_k . Esto nos conduce a las ecuaciones del método de Galerkin

$$\int_a^b \left[L \left(\sum_{k=1}^n a_k \omega_k \right) - f \right] \omega_i dx = 0 \quad (i = 1, 2, \dots, n).$$

Después de hallar los a_k mediante estas ecuaciones, construimos una expresión aproximada de la función (5).

Una fórmula similar se utiliza para la resolución de problemas variacionales por el método de Ritz, en análisis armónico aproximado de funciones y en otros muchos problemas.

Veamos otro ejemplo de simplificación de una ecuación. Supongamos que queremos hallar una función y de una o varias variables resolviendo cierta ecuación funcional; por ejemplo, una ecuación diferencial o integral. Como parámetros que caracterizan la función y elijamos sus valores $y_1, y_2, \dots, y_n$ en algún sistema de puntos.

La ecuación funcional debemos cambiarla entonces por un sistema de ecuaciones numéricas que contengan n incógnitas y_k ($k = 1, \dots, n$). Tal sustitución se puede hacer, por regla general, de varias maneras. Aquí siempre es necesario que nos esforcemos por que la solución del sistema numérico difiera suficientemente poco de la solución de la ecuación funcional.

A continuación damos varios ejemplos de este tipo de sustitución. Cuando resolvemos una ecuación diferencial de primer orden $y' = f(x, y)$ por el método de Euler, sustituimos esta ecuación por un sistema numérico recurrente que nos permita efectuar un cálculo aproximado de cada valor sucesivo de la función incógnita a partir del anterior (capítulo 5, §5):

$$y_{n+1} = y_n + (x_{n+1} - x_n)f(x_n, y_n).$$

Para una solución aproximada de la ecuación de Laplace

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

por el método de redes, sustituimos esta ecuación por un sistema lineal algebraico (capítulo 6, §5)

$$u(x+h, y) + u(x, y+h) + u(x-h, y) + u(x, y-h) - 4u(x, y) = 0.$$

Consideremos un ejemplo más de esta clase. Supongamos que queremos resolver numéricamente la ecuación integral

$$y(x) = f(x) + \int_a^b K(x, s) y(s) ds. \quad (6)$$

Los puntos en los que deseamos hallar los valores de la función incógnita $y(x)$ los denotaremos por $x_1, x_2, \dots, x_n$. Para formar el sistema de ecuaciones numéricas que sustituyan a (6), exigimos que la ecuación (6) se verifique no para todos los x del intervalo $a \leq x \leq b$, sino solamente en los puntos x_i ($i = 1, 2, \dots, n$)

$$y(x_i) = f(x_i) + \int_a^b K(x_i, s) y(s) ds.$$

Luego reemplazamos la integral por una cuadratura aproximada (por la regla trapezoidal, la regla de Simpson o cualquier otra)¹ con los puntos de división $x_1, \dots, x_n$

$$\int_a^b K(x_i, s) y(s) ds \approx \sum_{j=1}^n A_{ij} K(x_i, x_j) y(x_j).$$

Para determinar los valores deseados de $y(x_i)$, disponemos del sistema de ecuaciones algebraicas lineales

$$y(x_i) = f(x_i) + \sum_{j=1}^n A_{ij} K(x_i, x_j) y(x_j) \quad (i = 1, 2, \dots, n). \quad (7)$$

Observemos que todos los métodos considerados de buscar una función desconocida incluyen ciertos parámetros que la determinan aproximadamente. De este modo, la exactitud de estos métodos depende de cómo esté determinada la función por este sistema de parámetros; por ejemplo, de cómo se pueda aproximar por una expresión de la forma (7) o representar por sus valores en un cierto sistema de puntos. Las cuestiones de este tipo constituyen una rama particular de las matemáticas, llamada teoría de aproximación de funciones (capítulo 12). Esto demuestra el gran valor de la teoría de aproximación en la matemática aplicada.

Convergencia de los métodos de aproximación y una estimación del error

Examinemos con más detalle los requisitos para un método de cálculo. El más simple y fundamental de estos requisitos es la posibilidad de hallar la cantidad deseada con un grado prefijado de exactitud.

La exactitud requerida varía mucho de un problema a otro. Para

¹ Véase capítulo 12, §3.

ciertos cálculos técnicos groseros, dos o tres cifras decimales serán suficientes. La mayoría de los cálculos de ingeniería se efectúan con tres o cuatro cifras decimales. Pero la exactitud que se necesita en los cálculos científicos es mucho mayor. En líneas generales, la necesidad de exactitud ha ido aumentando en el transcurso del tiempo.

Particularmente importante, por consiguiente, son los métodos de aproximación y procesos que permiten obtener resultados con tanta exactitud como se desee. Tales métodos se denominan *convergentes*. Puesto que se encuentran con mucha frecuencia en la práctica y los requisitos que deben cumplir son característicos, los tendremos en cuenta en lo que sigue.

Sea x el valor exacto de una cantidad deseada. Para cada método de esta clase podemos construir una sucesión de aproximaciones $x_1, x_2, \dots, x_n, \dots$ a la solución x .

Después de indicar cómo se construyen las aproximaciones, el primer problema que se plantea es el de establecer la convergencia de las aproximaciones a la solución $x_n \rightarrow x$; y si el método no es siempre convergente, señalar las condiciones bajo las cuales lo es.

Una vez establecida la convergencia, surge el problema, más difícil y sutil, de estimar la *rapidez de convergencia*, es decir, estimar con qué rapidez converge x_n a la solución x para $n \rightarrow \infty$. Teóricamente, todo método convergente garantiza la posibilidad de hallar la solución con cualquier grado deseado de exactitud, si tomamos una aproximación x_n con índice suficientemente alto n . Pero, por regla general, cuanto mayor es n más complicado es calcular x_n . Así, si x_n converge lentamente a x , entonces obtener la exactitud necesaria puede llegar a exigir cálculos enormes.

En las mismas matemáticas, y especialmente en sus aplicaciones, se conocen muchos casos de procesos convergentes para hallar la solución x que necesitarían más trabajo de cálculo del que podrían efectuar hoy día las computadoras más rápidas².

² Mencionemos algunos ejemplos sencillos de procesos de cálculo que convergen lentamente. Es sabido que la serie

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

converge al logaritmo natural del número 2. Podemos hallar $\ln 2$ aproximadamente mediante esta serie, calculando la suma

$$s_n = \frac{1}{1} - \frac{1}{2} + \dots \pm \frac{1}{n}$$

de los n primeros términos, para n suficientemente grande. Pero se puede probar

La convergencia insuficientemente rápida es uno de los criterios por los que se juzga la desventaja de un método dado. Pero este criterio no es, naturalmente, el único, y al comparar distintos métodos debemos considerar otros muchos aspectos de la cuestión, en particular el de la conveniencia de efectuar los cálculos con máquinas. Entre dos métodos, a veces se prefiere utilizar aquel cuya convergencia es algo más lenta, si con él los cálculos son más fáciles de efectuar en una computadora.

El error cometido al reemplazar x por su valor aproximado x_n es igual a la diferencia $x - x_n$. Su valor exacto es desconocido, y para estimar la rapidez de convergencia, debemos hallar una cota superior para el valor absoluto de esta diferencia, es decir, una cantidad A_n tal que

$$|x - x_n| \leq A_n,$$

que se denomina *cota de error*. (Después daremos ejemplos de cotas A_n .) En consecuencia, el método usual de juzgar la rapidez de convergencia de un método es examinar la rapidez con que disminuye A_n cuando crece n . Para que la estimación refleje el grado efectivo de proximidad de x_n a x , es necesario que A_n difiera poco de $|x - x_n|$. Por otro lado, la cota A_n debe ser efectiva, es decir, tal que se pueda hallar, ya que de otra forma sería inservible.

Sea x una variable numérica cuyo valor deseamos determinar mediante cierta ecuación. Supongamos que nuestra ecuación se reduce a la forma

$$x = \phi(x). \quad (8)$$

A esta ecuación le aplicamos el *método de iteración*, que también se denomina a menudo *método de aproximaciones sucesivas*. Para exponer el método y las estimaciones con él relacionadas, examinaremos

que para calcular $\ln 2$ con un error menor que la mitad de la quinta cifra significativa, debemos tomar más de 100.000 términos de la serie. Hallar la suma de tal cantidad de términos, si utilizamos, por ejemplo, solamente una calculadora de mesa, sería muy laborioso. Otro ejemplo familiar es la serie

$$\frac{1}{\sqrt{2}} = 1 - \frac{1}{2 \cdot 1!} + \frac{1 \cdot 3}{2^2 \cdot 2!} - \frac{1 \cdot 3 \cdot 5}{2^3 \cdot 3!} + \frac{1 \cdot 3 \cdot 5 \cdot 7}{2^4 \cdot 4!} - \dots$$

Su convergencia es tan lenta que para calcular $1/\sqrt{2}$ con un error menor que 10^{-5} , necesitaríamos tomar aproximadamente 10^{10} términos, lo cual es difícil incluso con máquinas ultrarrápidas.

el caso de una ecuación numérica, aunque el método también se aplica a sistemas de ecuaciones numéricas, ecuaciones diferenciales, ecuaciones integrales y otros muchos casos. La aplicación del método a ecuaciones diferenciales ordinarias ya se ha ilustrado en el capítulo 5, §5.

Supondremos que hemos hallado de algún modo un valor aproximado x_0 para una raíz de la ecuación. Si x_0 fuera una solución exacta de la ecuación (8), después de sustituirla en su segundo miembro obtendríamos un resultado igual a x_0 . Pero como x_0 , por lo general, no es una solución exacta, el resultado de la sustitución será diferente de x_0 . Denotémoslo por $x_1 = \phi(x_0)$.

Para establecer en qué casos x_1 estará más próximo a la solución exacta que x_0 , recurrimos a una interpretación geométrica de nuestro problema. Consideremos la función

$$y = \phi(x). \quad (9)$$

Tomamos un eje numérico y representamos los números x e y mediante puntos del mismo. La ecuación (9) asigna a cada punto x un punto correspondiente y del mismo eje. Se puede considerar como una transformación puntual del eje numérico en sí mismo.

Tomemos el segmento $[x_1, x_2]$ del eje en cuestión. Mediante la transformación (9) los puntos x_1 y x_2 se transforman en los puntos

$$y_1 = \phi(x_1), \quad y_2 = \phi(x_2).$$

El segmento $[x_1, x_2]$ se transforma en el segmento $[y_1, y_2]$. La razón

$$k = \frac{|y_2 - y_1|}{|x_2 - x_1|}$$

se denomina «coeficiente de dilatación» del segmento bajo la transformación. Si $k < 1$, tendremos una contracción del segmento.

Volvamos a la ecuación (8). Esta nos dice que el punto x buscado debe transformarse en sí mismo bajo la transformación (9). Por tanto, resolver la ecuación (8) es equivalente a hallar el punto del eje que se transforma en sí mismo mediante (9), es decir, el punto que permanece invariante.

Consideremos ahora el segmento $[x, x_0]$, uno de cuyos extremos

es el punto fijo x , y el otro, el punto x_0 . Bajo la transformación dada, x_0 pasa a x_1 y el segmento $[x, x_0]$ al segmento $[x, x_1]$. Si la función ϕ tiene la propiedad de que bajo la transformación (9) todo segmento se contrae, entonces ciertamente x_1 estará más próximo que x_0 a la raíz de la ecuación (8).

Puesto que deseamos obtener aproximaciones que converjan a la solución exacta de (8), efectuamos la transformación repetidas veces y construimos la sucesión de números

$$x_1 = \phi(x_0), \quad x_2 = \phi(x_1), \quad \dots, \quad x_{n+1} = \phi(x_n), \quad \dots \quad (10)$$

Demostremos ahora que esta sucesión de aproximaciones converge³.

Supongamos que la función $\phi(x)$ está definida sobre cierto segmento $[a, b]$ y que la ecuación (9) define una transformación de $[a, b]$ en sí mismo, es decir, para cada x perteneciente a $[a, b]$, $y = \phi(x)$ también pertenece a $[a, b]$. Supondremos también que la aproximación inicial x_0 está en $[a, b]$; todas las aproximaciones sucesivas (10) estarán también en $[a, b]$. Bajo estas condiciones es cierto el siguiente teorema. Si $\phi(x)$ tiene una derivada ϕ' que verifica la condición

$$|\phi'| \leq q < 1$$

en $[a, b]$, entonces la siguiente proposición es cierta: la ecuación (8) tiene una raíz x^* en el segmento $[a, b]$. La sucesión (10) converge a esta raíz, y la rapidez de convergencia está caracterizada por la estimación

$$|x^* - x_n| \leq \frac{m}{1 - q} q^n,$$

donde $m = |x_0 - \phi(x_0)| = |x_0 - x_1|$. La ecuación (8) tiene una raíz única en $[a, b]$.

Para probar estas afirmaciones, estimamos la diferencia $|x_2 - x_1|$. Si la fórmula de Taylor es aplicable —capítulo 2, §9, (26)—, obtenemos, para $n = 0$

$$x_2 - x_1 = \phi(x_1) - \phi(x_0) = \phi'(\xi_0)(x_1 - x_0).$$

³ Debido a su interpretación geométrica, este teorema y otros como él se suelen llamar teoremas de contracción.

Entonces ξ_0 está entre x_1 y x_0 y por tanto pertenece al segmento $[a, b]$. De este modo $|\phi'(\xi_0)| \leq q$ y

$$|x_2 - x_1| \leq q |x_1 - x_0| = mq.$$

Análogamente

$$|x_3 - x_2| = |\phi(x_2) - \phi(x_1)| = |\phi'(\xi_1)(x_2 - x_1)| \leq q |x_2 - x_1| \leq mq^2.$$

Continuando estas estimaciones, tenemos, para cada valor de n , la desigualdad

$$|x_{n+1} - x_n| \leq mq^n. \quad (11)$$

Establezcamos ahora la convergencia de la sucesión x_n . A este fin consideramos la serie auxiliar

$$x_0 + (x_1 - x_0) + (x_2 - x_1) + \cdots + (x_n - x_{n-1}) + \cdots. \quad (12)$$

La suma parcial de los $n + 1$ primeros términos es igual a

$$s_{n+1} = x_0 + (x_1 - x_0) + \cdots + (x_n - x_{n-1}) = x_n.$$

Así, $\lim_{n \rightarrow \infty} s_{n+1} = \lim_{n \rightarrow \infty} x_n$, y la existencia de un límite finito para x_n es equivalente a la convergencia de la serie (12). Comparamos la serie (12) con la serie

$$|x_0| + m + mq + \cdots + mq^{n-1} + \cdots.$$

En virtud de la estimación (11), los términos de la serie (12) no son mayores en valor absoluto que los términos correspondientes de esta última. Pero esta serie, exceptuado su primer término $|x_0|$, es una progresión geométrica de razón q , y como $q < 1$, converge. La serie (12) es, por tanto, convergente también, y la sucesión (10) tiene un límite finito x^*

$$\lim_{n \rightarrow \infty} x_n = x^*.$$

Evidentemente, x^* pertenece al segmento $[a, b]$, ya que todos los x_n están en él.

Si en la ecuación $x_{n+1} = \phi(x_n)$ tomamos límites para $n \rightarrow \infty$, entonces obtenemos $x^* = \phi(x^*)$, que prueba que x^* verifica efecti-

vamente la ecuación (8). Estimemos ahora la distancia de x_n a x^* . Tomamos x_n y una aproximación posterior cualquiera x_{n+p}

$$\begin{aligned} |x_{n+p} - x_n| &= |(x_{n+p} - x_{n+p-1}) + (x_{n+p-1} - x_{n+p-2}) + \cdots + \\ &\quad + (x_{n+1} - x_n)| \leq mq^{n+p-1} + mq^{n+p-2} + \cdots + mq^n = \\ &= \frac{mq^n - mq^{p+n}}{1 - q}. \end{aligned}$$

Luego, para $p \rightarrow \infty$, en virtud de $x_{n+p} \rightarrow x^*$ y $q^{n+p} \rightarrow 0$, se deduce que

$$|x^* - x_n| \leq \frac{m}{1 - q} q^n.$$

Falta probar la afirmación de unicidad. Sea x' una solución cualquiera de la ecuación sobre $[a, b]$. Estimemos la diferencia $x' - x^*$:

$$|x' - x^*| = |\phi(x') - \phi(x^*)| = |\phi'(\xi)(x' - x^*)| \leq q |x' - x^*|,$$

de donde

$$(1 - q) |x' - x^*| \leq 0.$$

Como $1 - q > 0$, esta desigualdad sólo es posible para $|x' - x^*| = 0$, lo que significa que x' es idéntico a x^* .

El teorema no sólo da condiciones suficientes para la convergencia del método de iteración, sino que también nos permite estimar el número necesario de pasos implicados en el cálculo, es decir, cómo debemos tomar n de grande para obtener la exactitud requerida cuando se reemplaza x^* por x_n . Tal estimación es efectiva, ya que las cantidades m y q que aparecen en la desigualdad $|x^* - x_n| \leq (m/1 - q)q^n$ se pueden hallar estudiando la función ϕ .

Consideremos, por ejemplo, la ecuación $x = k \operatorname{tg} x$, que tiene muchas aplicaciones prácticas. Para precisar, consideremos el caso $k = 0,5$. Supongamos que queremos hallar la menor raíz positiva de la ecuación $x = (1/2) \operatorname{tg} x$. Debe ser próxima al punto 1 y algo mayor que 1, como puede establecerse fácilmente a la vista de una tabla o gráfica de la función $\operatorname{tg} x$.

Para garantizar la condición $|\phi'| \leq q < 1$, que forma parte del teorema de convergencia del método de iteración, invertimos la función $\operatorname{tg} x$ y consideramos la ecuación $x = \operatorname{arc} \operatorname{tg} 2x$ equivalente a la dada.

Damos aquí los resultados del cálculo. Para la aproximación original hemos tomado el valor $x_0 = 1$. Las aproximaciones siguien-

tes están calculadas mediante una tabla de la función $\text{arc tg } x$, de donde se obtienen los valores numéricos siguientes

$$\begin{aligned}x_1 &= \text{arc tg } 2 = 1,10715, \\x_2 &= \text{arc tg } 2,21430 = 1,14660, \\x_3 &= \text{arc tg } 2,29320 = 1,15959, \\x_4 &= \text{arc tg } 2,31918 = 1,16370, \\x_5 &= \text{arc tg } 2,32740 = 1,16498, \\x_6 &= \text{arc tg } 2,32996 = 1,16538, \\x_7 &= \text{arc tg } 2,33076 = 1,16550, \\x_8 &= \text{arc tg } 2,33100 = 1,16554, \\x_9 &= \text{arc tg } 2,33108 = 1,16555, \\x_{10} &= \text{arc tg } 2,33110 = 1,16556, \\x_{11} &= \text{arc tg } 2,33112 = 1,16556,\end{aligned}$$

Podemos detener aquí el cálculo, ya que en iteraciones posteriores se repite el valor de la raíz

$$x^* = 1,16556.$$

En la figura 1 se ilustran geoméricamente las aproximaciones de la raíz. Aquí x_n tiende tan rápidamente a x^* que x_4 ya no se puede distinguir de x^* en el diagrama.

Demos un ejemplo más del método de iteración. Vamos a resolver numéricamente la ecuación integral

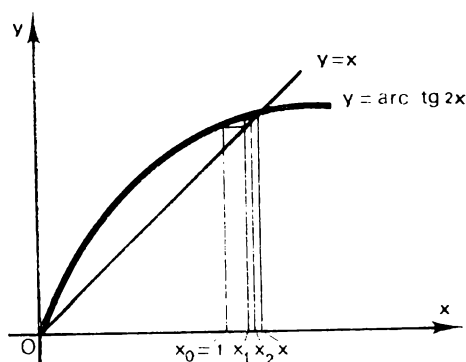


FIG. 1

$$y(x) = \frac{1}{6} \int_0^1 e^{xt} y(t) dt + e^x - \frac{1}{6} \frac{1}{x+1} (e^{x+1} - 1). \quad (13)$$

Su solución exacta es $y = e^x$.

Primero sustituimos la ecuación integral por un sistema de ecuaciones algebraicas lineales. Para ello dividimos el intervalo de inte-

gración $[0,1]$ en cuatro partes iguales mediante los puntos $t = 0, 1/4, 1/2, 3/4, 1$. Los valores de la función incógnita y en estos puntos los denotaremos por y_0, y_1, y_2, y_3, y_4 , respectivamente. Si imponemos que la ecuación se verifique para $x_0 = 0, 1/4, 1/2, 3/4, 1$, cuando la integral se reemplaza por la fórmula de Simpson para los cuatro intervalos parciales —capítulo 12, §3, (6)—, tenemos el siguiente sistema de ecuaciones y_k :

$$\begin{aligned}
 y_0 &= \frac{1}{6} (0,083333y_0 + 0,333333y_1 + 0,166667y_2 + \\
 &\quad + 0,333333y_3 + 0,083333y_4) + 0,713619, \\
 y_1 &= \frac{1}{6} (0,083333y_0 + 0,354831y_1 + 0,188858y_2 + \\
 &\quad + 0,402077y_3 + 0,107002y_4) + 0,951980, \\
 y_2 &= \frac{1}{6} (0,083333y_0 + 0,377716y_1 + 0,214004y_2 + \\
 &\quad + 0,484997y_3 + 0,137393y_4) + 1,261867, \\
 y_3 &= \frac{1}{6} (0,083333y_0 + 0,402077y_1 + 0,242499y_2 + \\
 &\quad + 0,585018y_3 + 0,176417y_4) + 1,664181, \\
 y_4 &= \frac{1}{6} (0,083333y_0 + 0,428008y_1 + 0,274787y_2 + \\
 &\quad + 0,705667y_3 + 0,226523y_4) + 2,185861.
 \end{aligned}$$

Este sistema se resuelve por el método de iteración. Como aproximación inicial a y_k ($k = 0, 1, 2, 3, 4$) tomaremos los términos constantes de las correspondientes ecuaciones: $y_0^{(0)} = 0,713619, y_1^{(0)} = 0,951980, \dots$. Los valores hallados en las aproximaciones sucesivas están dados en la tabla 1:

TABLA 1

Número de la aproximación	y_0	y_1	y_2	y_3	y_4
1	0,93428	1,20841	1,56129	2,01542	2,59972
2	0,98517	1,26699	1,62905	2,09419	2,69173
3	0,99667	1,28021	1,64433	2,11194	2,71245
4	0,99926	1,28319	1,64778	2,11595	2,71713
5	0,99985	1,28386	1,64856	2,11685	2,71818
6	0,99998	1,28402	1,64873	2,11705	2,71842
7	1,00001	1,28405	1,64877	2,11710	2,71847
Valor de la solución exacta	1,00000	1,28403	1,64872	2,11700	2,71828

Al final de la tabla se da el valor de la solución exacta para poder comparar. Aproximaciones ulteriores no mejorarían los valores de y_k . La divergencia en los últimos dígitos de las y_k proviene del error cometido al reemplazar la integral por la suma de Simpson.

Estabilidad de los métodos aproximados

Las necesidades del cálculo práctico imponen a los métodos de aproximación otro requisito general que se debe tener en cuenta debido a su gran importancia. Se trata de la *estabilidad* del proceso de cálculo. En esencia consiste en lo siguiente: cada método de aproximación conduce a un esquema de cálculo, y resulta que para obtener todos los números requeridos debemos efectuar una larga serie de pasos de acuerdo con el esquema. En cada paso el cálculo no se efectúa con exactitud, sino solamente con un número determinado de cifras significativas, y así en cada paso cometemos un pequeño error. Todos estos errores influirán en los resultados finales.

El esquema de cálculo adoptado puede resultar a veces muy poco satisfactorio, en el sentido de que los pequeños errores cometidos al principio van adquiriendo mayor influencia conforme se van efectuando los cálculos sucesivos, llegando a producir al final una gran desviación respecto a los valores exactos.

Consideremos la resolución numérica de una ecuación diferencial

$$y' = f(x, y)$$

con la condición inicial $y(x_0) = y_0$, en la que nos piden hallar los valores de $y(x)$ para valores equidistantes $x_k = x_0 + kh$ ($k = 0, 1, \dots$).

Supongamos efectuado el cálculo hasta el paso n , con los resultados indicados en la tabla 2.

TABLA 2

x	y	$y' = f$
x_0	y_0	y'_0
x_1	y_1	y'_1
$\dots$	$\dots$	$\dots$
x_{n-1}	y_{n-1}	y'_{n-1}
x_n	y_n	y'_n

Debemos hallar ahora y_{n+1} . Por el método de poligonales de Euler hacemos la aproximación

$$y_{n+1} = y_n + hy'_n. \quad (14)$$

Aquí y_{n+1} está calculada a partir de los números y_n e y'_n que aparecen en la última línea de la tabla 2. Supongamos que deseamos aumentar la exactitud y para ello utilizamos todas las cantidades que aparecen en las dos últimas líneas. Entonces podemos construir la fórmula

$$y_{n+1} = -4y_n + 5y_{n-1} + h(4y'_n + 2y'_{n-1}). \quad (15)$$

Observemos que si el cálculo es absolutamente exacto, es decir, con un número infinito de cifras significativas, entonces la fórmula (14) dará el resultado exacto siempre que y sea un polinomio lineal, y la fórmula (15) será exacta para todo polinomio de grado no superior a tres. A primera vista puede parecer que los resultados obtenidos aplicando la fórmula (15) deben ser más exactos que los hallados por el método de las poligonales. Sin embargo se puede comprobar fácilmente que la fórmula (15) es poco apropiada para el cálculo, ya que su aplicación puede producir un rápido aumento del error.

Los valores de la derivada y'_n e y'_{n-1} están multiplicados por un factor h pequeño, por lo que los errores en estos valores tienen menos influencia que los errores en y_n e y_{n-1} . Supongamos, para simplificar, que los valores de y' están calculados exactamente, con lo cual no hace falta tenerlos en cuenta para calcular el error en los dos casos anteriores. Supongamos que al hallar y_{n-1} cometemos un error de $+\epsilon$, y al hallar y_n un error de $-\epsilon$. Entonces, como muestra la ecuación (15), en y_{n+1} el error cometido es del orden de $+9\epsilon$. En y_{n+2} , de -41ϵ , y así irá aumentando sucesivamente. La fórmula (15) conduce a un proceso de cálculo que es inestable respecto a los errores; debemos, pues, descartarlo.

El ejemplo dado muestra hasta qué punto se pueden falsear los resultados por un esquema de cálculo inestable. Aquí hemos resuelto la ecuación $y' = y$ con la condición inicial $y_0 = 1$. La solución exacta es $y = e^x$. Para la resolución numérica hemos tomado valores equidistantes de la variable independiente x separados una distancia $h = 0,01$, es decir, $x_k = 0,01 k$. Hemos calculado una solución aproximada de dos formas: por el método de poligonales (14) y por la fórmula (15). A efectos de comparación, la tabla 3 da el valor de la solución exacta con siete cifras decimales.

Los valores aproximados de la solución hallada mediante la fór-

mula (15) son más exactos, al principio, que los obtenidos por el método de las poligonales, pero la inestabilidad de la fórmula (15) comienza en seguida a falsear los valores aproximados de y_k y conduce a números muy distintos de los verdaderos valores de y_k .

TABLA 3

x	Valores de la solución exacta	Valores de las soluciones aproximadas calculadas	
		por la fórmula (14)	por la fórmula (15)
0,00	1,0000000	1,0000000	1,0000000
0,01	1,0100502	1,0100000	1,0100502
0,02	1,0202013	1,0201000	1,0202012
0,03	1,0304545	1,0303010	1,0304553
0,04	1,0408108	1,0406040	1,0408070
0,05	1,0512711	1,0510100	1,0512899
0,06	1,0618365	1,0615201	1,0617431
0,07	1,0725082	1,0721353	1,0729726
0,08	1,0832871	1,0828567	1,0809789
0,09	1,0941743	1,0936853	1,1056460
0,10	1,1051709	1,1046222	1,0481559
0,11	1,1162781	1,1156684	1,3996456
0,12	1,1274969	1,1268250	-0,2808540

Elección de los métodos de cálculo

Todo cálculo se puede reducir, en resumidas cuentas, a las cuatro operaciones aritméticas de adición, sustracción, multiplicación y división. Describir un método de cálculo consiste en fijar los datos iniciales de los que se parte y después señalar qué operaciones aritméticas y en qué orden se han de efectuar para obtener los resultados deseados. Mostremos ahora con un sencillo ejemplo lo mucho que influyen en la organización de los cálculos la experiencia y conocimientos del matemático encargado de hacer el esquema y los excelentes resultados que se pueden obtener con una adecuada selección de métodos especialmente apropiados a la situación.

Queremos resolver el sistema de n ecuaciones con n incógnitas $x_1, x_2, \dots, x_n$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2,$$

$$\dots \dots \dots$$

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n.$$

En virtud de la teoría de sistemas algebraicos (capítulo 15, §3) disponemos de una expresión explícita para los valores de las incógnitas mediante determinantes:

$$x_j = \frac{\Delta_j}{\Delta} \quad (j = 1, 2, \dots, n). \quad (16)$$

Aquí Δ es el determinante del sistema

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix},$$

y Δ_j es el determinante obtenido de Δ reemplazando la columna j -ésima por la columna de los términos independientes del sistema.

Supongamos que deseamos utilizar la fórmula (16) para resolver el sistema y que hemos comenzado a calcular los determinantes aplicando la definición usual, sin recurrir a ninguna simplificación. ¿Cuántas multiplicaciones y divisiones serán necesarias? (Las adiciones y sustracciones no se tienen en cuenta, ya que son operaciones relativamente sencillas.) Nos encontramos ante el panorama de calcular $n + 1$ determinantes de orden n . Cada uno de ellos consta de $n!$ términos; cada término es el producto de n factores y en consecuencia requiere $n - 1$ multiplicaciones. Para calcular todos los determinantes debemos efectuar $(n + 1)n! \times (n - 1)$ multiplicaciones. El número total de multiplicaciones y divisiones será igual a $(n^2 - 1)n! + n$.

Consideremos ahora otro método para resolver el sistema: el de eliminación sucesiva de incógnitas. El esquema de cálculo que corresponde a este método va asociado al nombre de Gauss. Despejamos x_1 de la primera ecuación del sistema

$$x_1 = \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2 - \cdots - \frac{a_{1n}}{a_{11}} x_n.$$

Para ello necesitamos n divisiones. Sustituir x_1 en cada una de las $n - 1$ siguientes ecuaciones requiere n multiplicaciones. La eliminación de x_1 y la formación de $n - 1$ ecuaciones con las incógnitas $x_2, \dots, x_n$ requerirá entonces n^2 multiplicaciones y divisiones. Continuando este proceso, obtenemos que para calcular todos los valo-

res de x_j ($j = 1, \dots, n$) el método de eliminación requiere $n/6(2n^2 + 9n - 5)$ multiplicaciones y divisiones. Comparemos estos dos resultados. Para resolver un sistema de cinco ecuaciones en el primer caso necesitaríamos 2.885 multiplicaciones y divisiones, y en el segundo caso, 75.

Para un sistema de diez ecuaciones el número de operaciones será de $(10^3 - 1) 10! + 10 \approx 360.000.000$ y $10/6(2 \cdot 10^2 + 9 \cdot 10 - 5) = 475$, respectivamente. Vemos, pues, que la laboriosidad del cálculo depende enormemente del método elegido. Al organizar el esquema de cálculo es a menudo posible reducir muchísimo el trabajo mediante una selección racional del método.

§2. Los medios auxiliares más simples del cálculo

Tablas

Los medios auxiliares del cálculo más antiguos son las tablas. Las tablas más sencillas, es decir, la tabla de multiplicar y las de logaritmos o de funciones trigonométricas, son muy conocidas del lector. La gama de problemas que son resolubles en cuestiones prácticas aumenta continuamente. Los nuevos problemas se resuelven a menudo mediante la aplicación de nuevas fórmulas o pueden conducir a nuevas funciones, por lo que el número de tablas necesarias no cesa de aumentar.

Toda tabla, prescindiendo de cómo ha sido construida, contiene los resultados de cálculos anteriores y por consiguiente representa una especie de memoria matemática. Las tablas impresas o escritas están destinadas a ser leídas por los seres humanos. Pero también podríamos considerar tablas formadas de alguna manera especial, por ejemplo mediante agujeros perforados en tarjetas de una forma determinada, que estén destinadas a ser leídas por máquinas computadoras. Pero tales tablas son muy raras y no las discutiremos aquí.

Las tablas más utilizadas son las de valores de funciones. Si una función y depende solamente de una variable x , entonces la tabla más sencilla tiene la forma siguiente:

x	y
x_1	y_1
x_2	y_2
$\dots$	$\dots$
x_n	y_n

(17)

Las tablas de este tipo se denominan tablas de una entrada ⁴. De un modo directo sólo se pueden leer los valores correspondientes a los valores tabulados de x . Los que corresponden a valores de x que no figuren en la tabla es preciso hallarlos por interpolación de distintas clases, como se ha visto en el capítulo 12 ⁵. Por consiguiente, las tablas contienen a menudo, además de los valores de las funciones, ciertas cantidades auxiliares que hacen más fácil la interpolación. Normalmente suelen ser valores de la primera o segunda diferencia. Las tablas más especializadas requieren fórmulas de interpolación especiales, para las que se incluyen los datos correspondientes.

Para una función de dos variables $u = f(x, y)$, los valores de la función se disponen en una tabla de doble entrada de la siguiente forma:

$x \backslash y$	y_1	y_2		y_m
x_1	u_{11}	u_{12}		u_{1m}
x_2	u_{21}	u_{22}		u_{2m}
$\dots$	$\dots$	$\dots$		$\dots$
x_n	u_{n1}	u_{n2}		u_{nm}

(17')

Cada columna de la tabla es a su vez una tabla de una entrada, por lo que (17') es una colección de tablas de la forma (17). El tamaño de una tabla de una función de dos variables es, por regla general, mucho mayor que para una función de una variable con el mismo recorrido de las variables independientes. En vista de ello, las funciones de dos variables se tabulan con mucha menor frecuencia que las de una sola.

Para darnos una idea de cómo puede aumentar el tamaño de una

⁴ La columna puede ser muy larga y cabe, por tanto, fragmentarla en varias columnas más pequeñas por razones de comodidad. Pero, naturalmente, seguirá siendo una tabla de una entrada.

⁵ La interpolación, por regla general, es tanto más complicada cuanto más separados se hallen los valores tabulados x_i . Por otro lado, la exigencia relativa a la rapidez de interpolación puede variar ampliamente. En las tablas destinadas al uso de la artillería, la interpolación debe hacerse casi instantáneamente, «a primera vista». Pero en las tablas de precisión superior, destinadas para el uso científico, podemos permitirnos interpolación que requieran una serie completa de operaciones.

tabla al crecer el número de variables, consideremos el siguiente ejemplo. Supongamos que se quiere tabular una función de cuatro variables $f(x, y, z, t)$ para 100 valores de cada una de las variables. Supongamos que no hace falta calcular la función con mucha exactitud, sino sólo con tres cifras significativas. Si bajo tales condiciones tabulamos una función de una variable, la tabla completa de valores consta de cien números de tres dígitos, que se podrían escribir fácilmente en una página. Pero en una tabla de cuatro entradas para la función $f(x, y, z, t)$ tendremos 100^4 combinaciones de los valores de x, y, z, t y otros tantos valores de f , de lo cual es fácil calcular que la tabla ocuparía más de 300 volúmenes.

Debido a que tales tablas son tan poco manejables, las funciones de muchas variables rara vez se tabulan y solamente se hace en casos particularmente sencillos. En los últimos años se ha iniciado un estudio sistemático de clases de funciones de muchas variables cuyas tablas se pueden formar con un número de entradas menor que el número de variables. Al mismo tiempo se han iniciado estudios sobre la construcción más sencilla posible de tales tablas.

Veamos un ejemplo sencillo de una función de este tipo.

Queremos tabular la función u de tres variables x, y, z con la siguiente estructura:

$$u = f[\phi(x, y), z].$$

Está perfectamente claro que aquí nos podemos limitar a dos tablas de doble entrada si introducimos la variable auxiliar $t = \phi(x, y)$ y consideramos u como la función compuesta

$$u = f(t, z),$$

$$t = \phi(x, y).$$

Por conveniencia en el uso de estas tablas, podemos combinarlas de la manera siguiente. Consideramos la función $t = \phi(x, y)$ y despejamos y :

$$y = \Phi(x, t).$$

Teóricamente no existe ninguna diferencia entre tabular cualquiera de las funciones $t = \phi(x, y)$ o $y = \Phi(x, t)$, pero para nosotros es más conveniente tabular la segunda de ellas. Construimos dos tablas

de doble entrada para las funciones $y = \Phi(x, t)$ y $u = f(t, z)$ y las combinamos de la forma indicada en la tabla 4.

TABLA 4

x_1	x_2	$\dots$	x_i	$\dots$	t	z_1	z_2	$\dots$	z_k	$\dots$
					t_1					
					t_2					
					$\vdots$					
					$\vdots$					
$\dots$	$\dots$	$\dots$	y_j	$\dots$	t_j	$\dots$	$\dots$	$\dots$	u_{jk}	$\dots$

El valor de u que corresponde a los valores x_i, y_j, z_k se obtiene como sigue: buscamos la columna encabezada por x_i y la recorremos hacia abajo hasta encontrar el valor y_j (o uno próximo a él). Ahora nos movemos horizontalmente hasta encontrar el valor correspondiente de t . Seguimos esta horizontal hasta llegar a la columna z_k donde encontramos el valor buscado $u = f(x_i, y_j, z_k)$.

En este ejemplo vemos que, en vez de hacer una tabla de triple entrada, nos podemos restringir a dos de doble entrada con una regla muy sencilla para operar con ellas.

El uso de los distintos métodos posibles de reducir tablas nos permite, en ciertos casos, disminuir el tamaño de las mismas a la décima, centésima e incluso milésima parte en comparación con las tablas cuyo número de entrada es igual al de variables independientes.

Calculadoras de mesa

Casi tan antiguos como las tablas son otros artificios de cálculo. Algunos de ellos fueron utilizados incluso en la antigua Grecia.

Los primeros modelos de máquinas calculadoras fueron construidos en el siglo XVII por Pascal, Moreland y Leibnitz. Desde entonces las máquinas se han transformado y perfeccionado repetidamente, siendo ya muy utilizadas a finales del siglo pasado y especialmente a principios del actual.

Aquí sólo consideraremos ciertos modelos de máquinas y analizaremos la posibilidad de acelerar los cálculos que efectúan. Comencemos por las pequeñas, las llamadas calculadoras de mesa universales. Todas ellas, independientemente de su construcción, están diseñadas para efectuar las cuatro operaciones aritméticas; la multi-

plicación y división se realizan mediante series repetidas de adiciones y sustracciones.

Un modelo primitivo típico de una máquina de esta especie es el aritmómetro de Odner. Para introducir un número en el mecanismo ajustable se mueve una palanca el número de muescas que corresponda a cada dígito. Para sumar, se introducen los sumandos en el mecanismo ajustable y luego, mediante un giro de la manivela, se transfieren al acumulador, donde se suman automáticamente al número que ya estaba allí. La sustracción se efectúa mediante un giro de la manivela en sentido opuesto. La multiplicación se efectúa colocando el multiplicando en el mecanismo ajustable y sumándolo repetidamente consigo mismo para cada dígito del multiplicador. Por ejemplo, multiplicar por 45 corresponde a cinco adiciones repetidas del multiplicando y luego cuatro adiciones repetidas del mismo número trasladado un lugar a la izquierda.

Para la división se coloca el dividendo en el acumulador y el cociente se halla por sustracción repetida del divisor, dígito a dígito. El resultado está determinado por el número de giros de la manivela que se requieren en cada dígito para eliminar el número del acumulador.

Hemos dado aquí esta breve descripción de los cálculos solamente para mostrar la dirección que han seguido los perfeccionamientos ulteriores en las calculadoras de mesa. Algunos de estos perfeccionamientos se han limitado a hacer más manejables las máquinas, pero sin cambiar el esquema básico de su construcción. Una mejora de este tipo es la incorporación de la electricidad, que acelera el mecanismo de la máquina y evita tener que girar la manivela.

Con el fin de acelerar y simplificar la entrada de números en el mecanismo, se introdujeron teclas. La entrada de los dígitos se efectúa, no girando la palanca un número específico de muescas, sino simplemente pulsando la tecla correspondiente. Con el tiempo se inventaron calculadoras en las que basta introducir el número con el que se desea efectuar una determinada operación y pulsar la tecla que indica cuál de las cuatro operaciones se debe realizar. La máquina funciona desde este momento sin intervención humana. El perfeccionamiento de las calculadoras de mesa trajo consigo también un notable aumento de la rapidez, de modo que en los últimos modelos el resultado de una multiplicación se obtiene un segundo después de pulsar las teclas. Una mayor aceleración en el mecanismo de estas máquinas es evidentemente inútil, puesto que el tiempo que tarda el operador en pulsar las teclas y registrar los resultados es mucho mayor.

Máquinas digitales (de tarjetas perforadas) y máquinas de relé

Las máquinas digitales se idearon para cálculos estadísticos y para usos financieros e industriales. Están diseñadas para efectuar un gran número de operaciones poco complicadas de la misma clase. Para cálculos técnicos y científicos su utilidad es menor, debido a que su «memoria» es muy pequeña y debido también a la limitada posibilidad de establecer programas de cálculo. A pesar de estas deficiencias, las máquinas digitales, hasta la aparición de las rápidas máquinas electrónicas, fueron muy utilizadas en cálculos complicados y a gran escala cuando el proceso completo se podía reducir a una sucesión razonablemente corta de operaciones para efectuar a escala masiva (por ejemplo, en la preparación de tablas).

Los números con los que opera la máquina digital se introducen mediante tarjetas perforadas (Fig. 2). Los dígitos y símbolos se registran en la tarjeta mediante perforaciones en lugares determinados. La tarjeta se introduce en la máquina a través de un sistema de escobillas. Una escobilla bajo la cual quede un agujero cierra un circuito eléctrico y pone en funcionamiento una determinada fase de la máquina.

Los diferentes tipos de máquinas digitales están diseñados para trabajar en equipos, cada uno de los cuales consta de, al menos, las siguientes máquinas:

Una *perforadora* que sirve para practicar los agujeros en las tarjetas. Posee un teclado accionado a mano y trabaja a la velocidad de una máquina de escribir.

Una *clasificadora* diseñada para disponer las tarjetas en el orden en que se han de introducir en la máquina calculadora. Su velocidad es de 450-650 tarjetas por minuto.

Una *perforadora reproductora* o *reproductora*, que traslada perforaciones de una tarjeta a otra, compara dos grupos de tarjetas y selecciona aquellas con perforaciones específicas. Su velocidad es aproximadamente de 100 tarjetas por minuto.

Un *tabulador*, que efectúa las operaciones de adición y sustracción y también imprime los resultados. Puede manejar de 6.000 a 9.000 tarjetas por hora.

Un *multiplicador*, que suma, resta y multiplica números. Los resultados vienen dados en forma de perforaciones en las tarjetas. Trabajando con números de 6 ó 7 dígitos puede efectuar 700-1.000 multiplicaciones por hora.

Las máquinas digitales trabajan más bien despacio. Como cálculo

[illegible]

FIG. 2

aproximado de la cantidad de trabajo que pueden desarrollar diremos que el equipo anterior puede reemplazar a 12-18 calculadoras de mesa. Los primeros intentos para crear máquinas más rápidas condujeron a la construcción de máquinas de relé basadas en la aplicación de relés electromecánicos. La rapidez de tales máquinas resulta ser unas diez veces mayor que la de la máquina digital. Las ventajas en otros aspectos fueron también notables: las máquinas de relé ejecutan programas de cálculo complicados y tienen un sistema de control flexible que aumenta enormemente la gama de problemas técnicos y científicos resolubles con máquinas. Sin embargo, la aparición de éstas coincidió con la creación de los primeros modelos de máquinas electrónicas con control programado, y esto llevó a un brusco aumento en la velocidad de trabajo. Como una indicación del gran aumento de velocidad que trajo consigo la invención de las máquinas electrónicas, podemos apuntar que el tiempo requerido para un cambio de estado en un tubo electrónico viene medido en millonésimas de segundo.

Máquinas matemáticas de acción continua (máquinas analógicas)

Las máquinas matemáticas de acción continua están compuestas de sistemas físicos (aparatos mecánicos, circuitos eléctricos, etc.), y contruidas de tal manera que las interrelaciones numéricas que se dan entre los parámetros continuamente variables del sistema (desplazamientos, ángulos de giro, corrientes, voltajes, etc.) son las mismas que las que se dan entre las magnitudes correspondientes del problema a resolver. Tales máquinas reciben a menudo el nombre de *analógicas*.

Cualquier máquina de acción continua está especialmente diseñada para un tipo muy especial de problemas.

La exactitud con que la máquina da la solución depende de la calidad de las partes componentes, del montaje y calibración de la máquina, de los errores inerciales en su funcionamiento, etc. Sobre la base de una larga experiencia en el uso de máquinas se ha establecido que, por regla general, son capaces de una exactitud de dos o tres cifras significativas. En este aspecto las máquinas analógicas son notablemente inferiores a las digitales, cuya exactitud es teóricamente ilimitada.

Una característica importante de las máquinas analógicas es que se prestan a la resolución de un gran número de problemas del mismo tipo. Además, a menudo llegan a la solución con una rapidez considerablemente mayor que una máquina digital. Su principal ventaja consiste en que en muchos casos es más conveniente introducir los

datos iniciales del problema en ellas; asimismo, los resultados se obtienen a menudo de una forma más conveniente.

Existen muchos tipos diferentes de máquinas de simulación. Es posible crear máquinas, o partes de máquinas, que sean modelos de distintas operaciones matemáticas: adición, multiplicación, integración, diferenciación, etc. Podemos asimismo simular distintas fórmulas utilizadas en cálculo; por ejemplo, podemos construir máquinas para calcular valores de polinomios a los coeficientes de Fourier en análisis armónico de funciones, o para simular ecuaciones numéricas o funcionales. Las muchas analogías que existen entre problemas de ramas totalmente distintas de la ciencia llevan a las mismas ecuaciones diferenciales. La identidad de estas ecuaciones nos permite, por ejemplo, simular fenómenos de calor por medios eléctricos y resolver problemas técnicos de calor mediante medidas eléctricas, procedimiento indudablemente útil, ya que las medidas eléctricas son más exactas que las de calor y mucho más fáciles de hacer.

A la vista de la gran cantidad de máquinas analógicas, es imposible dar en pocas palabras una descripción de ellas o incluso de los prin-

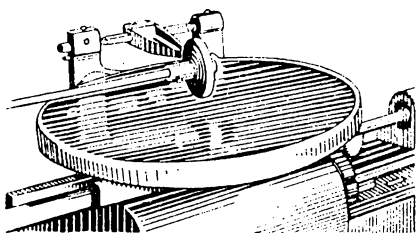


FIG. 3

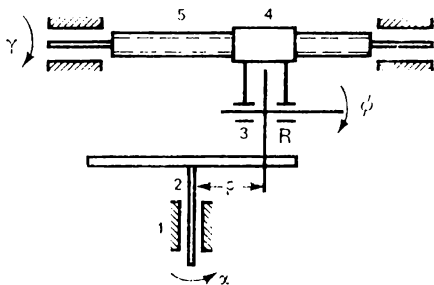


FIG. 4

cipios de su construcción. Para dar al lector por lo menos una idea de cómo se pueden simular problemas matemáticos, veamos una descripción breve de dos máquinas matemáticas sencillas, una de las cuales está diseñada para integración de funciones y la otra para la resolución aproximada de la ecuación de Laplace.

El *integrador de fricción* (Fig. 3) está diseñado, como el nombre indica, para integrar funciones. Trabaja por fricción. La idea básica de su construcción se indica en la figura 4, donde la componente 1 es la base del integrador, 2 es un disco de fricción horizontal con un

vástago vertical, 3 es una rueda de fricción, es decir, una rueda de borde liso que, además de rodar a lo largo del disco, puede también moverse en el plano perpendicular al de rodamiento. Las componentes 4 y 5 constituyen un mecanismo a rosca en el que el tornillo 4 está conectado al soporte de la rueda. Si el paso de rosca se denota por h , entonces un giro del tornillo en un ángulo γ desplazará a la rueda una distancia $\varrho = h\gamma$ en el plano del dibujo.

Supongamos que el vástago del disco ha girado un ángulo da . El punto de contacto de la rueda se moverá entonces un arco de longitud ϱda . Si la rueda se mueve sobre el disco sin deslizamiento, el ángulo de rotación de la misma será igual a

$$d\phi = \frac{\varrho}{R} da = \frac{h}{R} \gamma da.$$

Suponemos que la rotación del vástago del disco comenzó con un ángulo α_0 y el ángulo inicial de rotación de la rueda era ϕ_0 . De esta ecuación obtenemos por integración

$$\phi - \phi_0 = \frac{h}{R} \int_{\alpha_0}^{\alpha} \gamma da.$$

Mediante una elección conveniente de la relación entre los ángulos γ y α , podemos utilizar el integrador de fricción para calcular una integral en una gran variedad de casos. Mediante mecanismos integradores es posible obtener una solución mecánica de muchas ecuaciones diferenciales.

Veamos el segundo ejemplo. Sea un dominio Ω del plano, limitado por la curva l . Se desea hallar una función u que en el interior del dominio verifique la ecuación de Laplace

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

y sobre el contorno l tome valores prefijados

$$u|_l = f.$$

Introducimos una red cuadrada de puntos

$$x_k = x_0 + kh, \quad y_k = y_0 + kh, \quad k = 0, \pm 1, \pm 2, \dots,$$

y reemplazamos el dominio Ω por un polígono compuesto por cuadrados. En lugar del contorno l tenemos entonces una poligonal. Traslademos los valores de contorno de f en l a esta poligonal. El valor de la función incógnita u en un nudo (x_j, y_k) se denota por u_{jk} . Para obtener una solución aproximada de la ecuación de Laplace en Ω , la reemplazamos por un sistema algebraico, que se debe verificar para todos los puntos interiores del dominio:

$$u_{jk} = \frac{1}{4}(u_{j+1,k} + u_{j,k+1} + u_{j-1,k} + u_{j,k-1}).$$

Para resolver este sistema algebraico podemos construir el siguiente modelo eléctrico. Introducimos en el plano una red conductora bidimensional cuyo esquema se ilustra en la figura 5. La resistencia entre dos nudos se supone que es siempre la misma. En los nudos de la frontera de la red aplicamos voltajes iguales a los valores de contorno de u en ellos. Estos voltajes determinarán el voltaje de todos los puntos interiores de la red. Denotemos por $U_{j,k}$ el voltaje en el nudo (x_j, y_k) . Si aplicamos la ley de Kirchhoff al nudo (x_j, y_k) , está claro que en él se verificará la ecuación

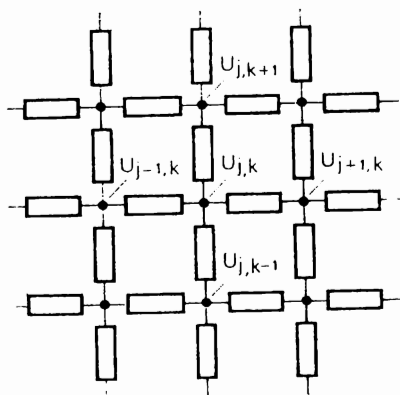


FIG. 5

$$\frac{1}{R} [(U_{j+1,k} - U_{j,k}) + (U_{j,k+1} - U_{j,k}) + (U_{j-1,k} - U_{j,k}) + (U_{j,k-1} - U_{j,k})] = 0,$$

que sólo difiere de la anterior ecuación de nuestro sistema algebraico en la notación. En los nudos de la red los valores u_{jk} de la solución del sistema algebraico deben coincidir con los voltajes U_{jk} , que se pueden obtener del modelo mediante medidas eléctricas normales.

INDICE ANALITICO

A

- Abel, N. H., I: 99, 318, 332, 333, 335; II: 171; III: 332, 343
- Abscisa, I: 260
- Adams, I., I: 21
- Adición, III: 26
- Aditiva, propiedad, II: 238
- Aditividad total de la medida, III: 43
- Ado, I. D., III: 396
- Afin, clasificación de curvas, figuras, superficies, etc., I: 283
- espacio, III: 77, 164
- propiedades de las figuras, I: 283
- Agregado; véase Conjunto
- Al-Biruni, I: 60
- Aleksandrov, A. D., I: 302; II: 140
- Aleksandrov, P. S., III: 41, 195, n. 3 p. 239, 260, 261, 265, 266, 373
- Alexander, J., III: 260, 261
- Algebra, I: 63-4, 82-3, 315-6; III: 309, 385
- abstracta, I: 319
- asociativa, III: 385
- axiomática, I: 319
- base de, III: 387
- con división, III: 394
- estructura de, III: 393
- topológica, III: 261
- ideal de, III: 392
- isomorfa, III: 388
- de Lie, III: 395
- lineal, III: 55
- de matrices, III: 390-391
- nihilpotente, III: 394
- rango de, III: 387
- semisimple, III: 394, 398
- simple, III: 394
- sistemas hipercomplejos, III: 310, 385
- teorema fundamental del, I: 338-52
- vectorial, I: 261; III: 381
- Algebraico(a)
- complemento, III: 87
- entero, II: 240
- operación, III: 406-407
- solución, I: 320
- variedad, III: 402
- Algebrización, I: 316

- Algoritmo, II: 238, 239
 Euclídeo, II: 238
 Al-Kharizmi, I: 61, 62, 63, 317
 Alternado, grupo, III: 327
 Análisis, I: 68, 73, 83, 91
 funcional, I: 84, 311-12; III: 265, 267
 Analítica, geometría, I: 68, 70, 225, 228-9
 en el espacio, I: 266-7
 infinito dimensional, I: 311
 prolongación, II: 222, 224, 225
 superficie, II: 125
 teoría de números, II: 240
 Analiticidad de una función compleja, II: 219
 Andronov, A. A., I: 396, 424; III: 254, 255
 Angulo, entre vectores, III: 77, 270
 de paralelismo, III: 133-134, 136, 138
 Anillo, III: 398
 de clases residuales, III: 400
 factor, III: 400
 no conmutativo, III: 403
 de operadores, III: 404
 principales, III: 401
 semisimple, III: 404
 sin divisores de cero, III: 401
 Apolonio, I: 57, 58, 70
 Aproximación, por defecto, I: 364, 366
 por exceso, I: 364, 367
 fórmula de, II: 356
 de funciones, II: 311
 de integrales definidas, II: 324
 métodos de, II: 355
 óptima, II: 313, 337
 óptima uniforme, II: 313, 330, 340
 por polinomios, II: 314, 333
 Arco, longitud de, II: 82
 Argumentos de un número complejo, I: 340
 Aristóteles, I: 92
 Aritmética, I: 24
 Armónica, función, II: 30
 Arquímedes, I: 33, 57, 58, 59, 71, 91, 97, 167; III: 31
 principio de, III: 31
 Aryabhata, I: n. 18 p. 60
 Asintótica, convergencia, III: 135
 Asintótica, ley, II: 246
 Asintótica en un punto, II: 77
 Asociados, elementos, III: 401
 Asociativa, álgebra, III: 385
 Asociatividad, III: 26
 Autofunción, II: 47; III: 295
 Autointersección, III: 238
 Automorfismo, III: 317
 Automorfismo, grupo de, III: 345
 Autooscilaciones u oscilaciones propias, I: 391
 Autovalor(es)
 espectro continuo de, III: 304
 de operadores, III: 301
 de una transformación lineal, III: 105
 Autovector(es), III: 104-105
 de operadores, III: 301-302
 de una transformación lineal, III: 104-105
 Autovibraciones, I: 391; III: 293
 Axioma(s), de adición, III: 26
 de continuidad, I: 347; III: 29, 153
 de distributividad, III: 27
 de un espacio lineal, III: 68
 de geometría, III: 151
 de incidencia, III: 152
 del movimiento, III: 153
 de multiplicación, III: 26
 de orden, III: 27, 152
 de paralelismo, III: 153
 de teoría de la probabilidad, II: 272
 Axiomática, álgebra, I: 319
- B**
- Banach, S., I: 84
 Base, de un álgebra, III: 387
 de un espacio, III: 71-72
 Básico, intervalo, II: 257
 Bernoulli, D., I: 74, 98, 369; III: 281
 Bernstein, S. N., II: 290, 307, 314, 329, 336, 340
 Bernstein, polinomio de, II: 339, 340
 Bertrand, J. L. F., II: 245

Bessel, F. W., I: 222
 Bézout, E., I: 318
 Bhaskara, I: n. 18 p. 60
 Birkhoff, G., I: 425; III: 260
 Blaske, W., II: 143
 Bogoljubov, N. N., I: 424, 425; II: 302
 Boltjanski, V. G., III: n. 12 p. 260, 261
 Bólyai, F., III: 127, 131, 375
 Bólyai, J., I: 80; III: 131
 Bolzano, B., I: 83, 347, 348; III: 16, 36
 Bonnet, P. O., II: 128
 Borel, E., I: 83; II: 314
 Boyle, K., I: 105, 293; II: 29; III: 174
 Brahe, T., I: 242
 Brahmagupta, I: n. 18 p. 60
 Braquistócrona, II: 145, 146
 Briggs, H., I: 64
 Brouwer, L. E. J., III: 259, 260
 Budan, I: 357
 Bunjakowski, V. Ja., III: 79, 80, 276, 277

C

Cadena, III: 248
 bidimensional, III: 249
 unidimensional, III: 248
 Caja canónica, III: 108
 Cálculo, diferencial, I: 71, 166
 integral, I: 71, 72, 166
 matricial, III: 117
 técnicas de II: 355
 Cálculo de variaciones, II: 145
 funciones admisibles, II: 152
 problema variacional en, II: 157, 159
 soluciones aproximadas en, II: 164, 165
 Calor, vector conducción de, II: 21
 Cambio de variable, I: 175-6
 Campo(s)
 algebraico, III: 27
 continuo de direcciones, III: 252
 de descomposición, III: 344
 de direcciones, I: 398; III: 252-254
 escalar, II: 16

vectorial, II: 15; III: 252
 continuo, III: 252
 Canónica, forma de una ecuación, I: 254
 de una matriz, III: 111-112
 forma cuadrática, III: 113
 matriz de Jordan, III: 108
 Cantor, G., I: n. 14 p. 47, 79, 83; III: 18, 29, 31-34, 37, 41, 259
 conjunto perfecto de, III: 39
 principio de continuidad de, III: 29
 Cantoriana, continuidad, III: 29
 de una figura, III: 194
 Caracteres, de grupos, III: 367
 Característica(o), ecuación, de una ecuación lineal homogénea, I: 389
 ecuación cuadrática, I: 292
 frecuencia; véase Autofunciones
 función; véase Autofunciones
 oscilaciones, I: 391
 Cardano, G., I: 318, 321, 322, 324, 325
 Cartan, E., III: 214, 358, 398
 Cauchy, A., I: 74, 83, 99, 117, 119, 171, 172; II: 180, 183, 187, 194, 198, 204; III: 16, 30, 31, 79, 80, 276, 277
 principio de, III: 30
 Cauchy-Riemann, ecuaciones de, II: 179, 206
 Cavalieri, B., I: 71
 Cayley, A., III: 169, 351
 Centro, I: 415
 Cero, I: 32
 Cerrado(a), clase de curvas o superficies, II: 139
 conjunto, III: 37, 262
 sistema ortogonal de funciones, III: 287
 superficie, III: 236, 244, 246
 superficie convexa, II: 140
 Cíclica, permutación, III: 328
 Cíclico, subgrupo, III: 331
 Ciclo(s), fundamental, III: 251-252
 límite, I: 421; III: 253
 límite estable, I: 422
 longitud de, III: 329-330
 permutación en, III: 328
 en topología, III: 250
 uni-dimensional, III: 249

- Ciclotómicas, ecuaciones, I: 332
 Cilíndricas, funciones, III: 282
 Cilindro elíptico, I: 270
 hiperbólico, I: 270
 parabólico, I: 270
 Circulación de flujo, II: 192
 Circulatorio, movimiento, II: 191
 Círculo de convergencia, II: 173
 Círculo límite, III: 139
 Circunferencia, no orientada, III: 337
 orientada, III: 336
 Clairaut, A. C., II: 178
 Clase cerrada de curvas y superficies, II: 139
 Clase, III: 352
 a la derecha, III: 352
 a la izquierda, III: 352
 Cociente por la derecha, III: 332
 Cociente por la izquierda, III: 33
 Cociente de diferencias de una función, II: 60
 Codazzi, II: 80, 128, 134
 ecuaciones de, II: 80, 128, 134
 Coeficiente de dilatación, II: 196
 363
 Coeficiente de una matriz, III: 62
 de transformación, III: 101
 Cohn-Vossen, S. E., III: n. 2 p. 343
 Cohomología, III: 261
 Colón, I: 91
 Columna, de una matriz, III: 59
 Combinatoria, topología, III: 260
 261
 Complejo(a), raíz, I: 327
 variable(s), ecuaciones de Cauchy-Riemann para, II: 179
 coeficiente de dilatación en, II: 196
 coordenadas, I: 312
 espacio, I: 312
 número, III: 374-375, II: 167-183
 argumento del, I: 340
 Complejo, plano, I: 352
 Complejos, enteros, II: 265, 267
 Completo, sistema de sucesos, II: 273
 Completitud de la recta numérica, III: 33
 composición de grupos, III: 347
 Computacional, matemática, I: 86
 Concavidad, I: 146-7
 Condición necesaria de convergencia, I: 209
 Conexión, III: 194
 Conexión de superficie, III: 235
 de variedades, III: 251
 Conforme, geometría, II: 143
 Conforme, transformación, II: 197, 200
 Congruencias, II: 137
 Congruencias de rectas, II: 137
 Congruentes, superficies, II: 128
 Cónicas, secciones, I: 252
 foco de, I: 248
 Conjunto(s), abierto, II: 37
 acotado, III: 33-34
 inferiormente, III: 34
 superiormente, III: 34
 acotado, e infinito de puntos, III: 36
 adherente, III: 194, 263
 aplicación unívoca de, III: 353
 cardinalidad de un, III: 22
 cero-dimensional, III: n. 11 p. 260
 cerrado, III: 37
 colineal de puntos, I: 298
 complemento de un, III: 38
 definición inductiva de dimensión de, III: n. 11 p. 260
 densidad de un, III: 27
 envoltura de, III: 42
 extremo inferior de, III: 34
 extremo superior de, III: 30-34
 finito, III: 20
 de funciones lineales, III: 57
 igual, III: 19
 infinito, III: 20, 36
 intersección de, III: 19
 lineal, III: 33
 linealmente ordenado, III: 404
 medible, III: 43
 n-dimensional, III: n. 11 p. 260
 no acotado, III: 33
 numerable, III: 23-24
 números racionales, III: 24
 ordenado, III: 33, 404
 parcialmente ordenado, III: 404
 perfecto, III: 39
 teoría de, III: 17

- teoría descriptiva de, III: 40
- totalmente ordenado, III: 404
- unión de, III: 19
- vacío, III: 19
- Conmutación, III: 396
- Conmutador, III: 340
 - de elementos de grupo, III: 333
- Conmutadores, subgrupo de, III: 333
- Conmutatividad, III: 26
- Conmutativo, grupo, III: 332
- Cono de segundo orden, I: 271
- Constante de estructura, III: 388
- Continuo(a), aplicación
 - función no regular, II: 323
 - funcional, III: 298
 - funciones, I: 117
 - grupo, III: 356, 358
 - grupo de transformaciones, III: 356
 - en un intervalo, I: 118
 - lateralmente, I: 119
 - en un punto, I: 118
 - representación, III: 358
 - transformación, III: 195, 232, 233, 263; II: 126
 - de medio continuos, I: 284-285; III: 257
 - variable, I: 54
 - no numerabilidad del, III: 32
 - número cardinal del, III: 25
- Contorno, condiciones de, II: 31, 34
 - condiciones homogéneas de, II: 36
 - problema de, II: 29, 33, 155
- Contracción del plano a una línea, I: 276, 285
- Contracción, teoremas de, II: 364
- Convergencia, asintótica, III: 135
 - círculo de, II: 173
 - dominio de, II: 171
 - «en media», III: 283
 - rapidez de, II: 361
 - de series, I: 209
- Convergente, onda esférica, II: 43
- Convergentes, métodos, II: 361
- Convergentes, series, I: 209
 - absolutamente, I: 211
- Convexa hacia abajo, I: 147
- Convexidad, I: 146-147
- Convexo, cuerpo, II: 140-142
- Coordenadas, curvilíneas, II: 114, 129
 - método de, I: 229
 - polares, I: 310
 - de un punto, I: 227
 - rectangulares, I: 260
 - sistema de, I: 281
 - transformación de, I: 255; III: 102-103
 - vectoriales, I: 263; III: 74
- Copérnico, N., I: 91, 225
- Correlación, coeficiente de, II: 288
- Correspondencia, uno a uno, III: 21
- Corriente, línea de, II: 185
- Cortadura de Dedekind, III: 29
- Corte en funciones complejas, II: 182
- Cota, I: 260
- Cota superior, III: 30
- Coulomb, C. A., I: 76
- Criba de Eratóstenes, I: 58; II: 241
- Cuadrática, ecuación, I: 321
 - característica, I: 292
- Cuaterno, norma de, III: 380
- Cuaternos, parte vectorial de, III: 380, 382-383
 - parte escalar de, III: 380
- Cuaternos conjugados, III: 378, 379, 380-381
- Cúbica, ecuación, I: 321
- Cubo tetradimensional, III: 177
- Cuerpo, III: 378
 - algebraico, III: 27
 - ordenado, III: 27
 - de base de álgebra, III: 344, 386
 - no simétrico, III: 319
 - simétrico, III: 318
- Curva(s), clases afines de, I: 283
 - clases cerradas de, II: 138
 - curvatura de, II: 87-90
 - ecuación de, I: 266
 - equidistante, III: 135
 - en el espacio, longitud del arco de, II: 83
 - familias de, II: 234
 - integral, I: 399; III, 253
 - irreducible, III: 403
 - longitud de, II: 83

matemática, II: 75
 n -dimensional, III: 184
 de nivel, I: 181, 267
 Curvatura, cero de una curva, II: 87, 90
 constante, III: 211
 del espacio, III: 209
 Gaussiana, II: 108
 media, II: 108
 principal, II: 103
 sobre una superficie, II: 101-108
 total; véase Curvatura Gaussiana
 Curvatura, tensor, III: 208
 Curvilínea(s), coordenadas, II: 114
 129
 integral, I: 201-203

CH

Chaplygin, S. A., II: 191, 193
 Chebotarev, H. G., I: 337, 338;
 III: 396, 402, 408
 Chebichev, P. L., I: 84, 87, 99,
 178; II: 79, 240, 243-246, 249,
 252, 263, 286-290, 329-334,
 335, 341, 350, 351; III: 282
 aproximación de, II: 313, 330-
 332, 340
 desigualdad de, II: 287
 polinomios de, II: 322, 323, 334
 Chech, E., II: 143
 Chudakov, N. G., II: 247

D

D'Alembert, J., I: 211; III: 127,
 281
 criterio de series convergentes,
 I: 211
 lema de, I: 349
 Dandelin, G. P., I: n. 14 p. 369
 Darboux, G., II: 80, 125
 Darwin, C. R., III: 131
 Da Vinci, L., I: 91; III: 160
 Dedekind, R., I: n. 14 p. 47, 83;
 II: 240; III: 29-32, 406
 Dedekind, continuidad de, I: 347;
 III: 29-30, 153
 cortadura, III: 28-29

principio de, III: 29
 retículo, III: 406
 Definida, integral, I: 165
 Deformabilidad de una superficie,
 II: 142
 Deformación de una superficie, II:
 115, 125, 142
 Deformación uniforme de una
 membrana, II: 147
 del Ferro, Scipio, I: 321, 329
 de Jonquières, III: 245
 Demócrito, I: 40, 45, 53, 54
 Densidad lineal, III: 292
 Dependencia, funcional, II: 149
 lineal, III: 69
 Derivación, reglas de, I: 132
 término a término, I: 212
 Derivada, de un cociente, I: 134
 de una función, I: 124
 de orden superior, I: 145
 parcial, I: 182
 parcial cruzada, I: 184
 de un producto, I: 133
 segunda, I: 146
 de una suma, I: 132
 Desargues G., III: 160
 Desarrollo,
 de una función en serie trigono-
 métrica, II: 243; III: 284
 en un sistema ortogonal de fun-
 ciones, III: 279, 282
 Descartes, R., I: n. 13 p. 46, 59,
 63, 68, 69, 70, 85, 96, 226-231,
 233-234, 236, 237, 260, 316, 317,
 318, 355-358; II: 129; III: 243
 Descartes, ley de los signos de, I:
 355
 Descomposición, de números ente-
 ros, II: 265
 de una permutación en ciclos,
 III: 328
 simplicial, III: 247
 de una variedad, III: 403
 Desviación
 típica, II: 287
 de frecuencia, II: 286
 probable, II: 276
 Determinados, procesos, II: 299
 Determinante(s), de los coeficien-
 tes de una matriz, II: 89

- complemento algebraico en, III: 87
- desarrollo de, III: 89
- de una matriz, III: 83
- menor, III: 87
- de orden n , III: 86-87
- Diagonal, proceso, III: 24
- Diámetro, de una región, I: n. 34 p. 200
- teoría de Newton de, I: 237
- Diferencia de dos conjuntos, III: 20
- Diferenciabilidad de una función, II: 177
- Diferenciable, función, I: 129; II: 195
- Diferencial(es), ecuación(es), I: 373 y sigs.
- campo de direcciones, I: 398; III: 253-254
- centro de, I: 415; III: 258
- en derivadas parciales, I: 376
- estable, I: 417
- estocástica, II: 307-309
- de Euler, II: 151-154
- foco en, I: 415
- lineal, I: 386-387
- lineal homogénea, I: 387
- orden de, I: 384
- ordinaria, I: 373, 376, 383
- «poco sensibles», I: 396
- solución general de, I: 385
- Diferencial de una función, I: 150
- de varias variables, I: 185
- Dilatación, coeficiente de, II: 196, 363
- Dimensión(es), del espacio, III: 71
- de una variedad, III: 242-243
- Dimensión, teoría de, III: 259
- Diofanto, I: 59, 60
- Dirección, números de, I: 407
- principal, II: 103
- Directos, métodos, I: 402
- Directriz de una parábola, I: 248
- Dirichlet, P. G. L., I: 108; II: 74; III: 46, 49
- función de, III: 46
- Discontinua, función, I: 119
- Discontinuidad débil, II: 68
- Discreto, grupo de movimientos, III: 338, 342
- Discreto, grupo simétrico, III: 338
- Disjuntas, clases, en grupos, III: 349
- Disjuntos, conjuntos, III: 20
- Distancia, en un espacio n -dimensional, III: 262
- entre funciones, III: 263-264
- entre puntos, III: 33
- Distribución de números primos, II: 242-244
- Distributiva, ley, III: 20
- Distributividad, III: 27
- Divergencia, teorema de Gauss de: véase Ostogradski, fórmula de
- Divergentes, series, I: 207
- Divisibilidad, II: 236
- División, álgebra con, III: 394
- Divisor, de cero, III: 401
- normal, III: 352
- de uno, III: 401-402
- Dominio, de convergencia, II: 171
- de definición, II: 149; III: 47, 299
- de una función, I: 103
- integral, III: 401
- de racionalidad, III: 344
- simplemente conexo, II: 201
- de valores, III: 191
- de variación de variables, III: 191
- Dualidad, teorema de, III: 261
- Durero, A., III: 160

E

- Ecuación(es), característica, homogénea lineal, I: 389
- de Cauchy-Riemann para variables complejas, II: 179, 206
- centro en, I: 415
- ciclotómica, I: 333
- de conservación de la energía calorífica, II: 22
- de continuidad, II: 20
- cuadrática, I: 321
- de cuarto grado, I: 324
- de una cuerda vibrante, II: 26
- de una curva, I: 266
- diferencial, I: 373-425

- construcción de soluciones, II: 43
- estocástica, II: 307-309
- estructuralmente estable, I: 417
- lineal, I: 386
- lineal homogénea, I: 387
- lineal no homogénea, I: 387, 392-395
- ordinaria, I: 373, 375, 384
- solución general de, I: 385
- forma canónica de, I: 255
- funcional, I: 374
- homogénea correspondiente, II: 43
- integral, III: 288-296
 - homogénea, III: 292
 - no homogénea, III: 294
 - de segunda especie, II: 57
- de Laplace, II: 30, 194
- de una línea recta, I: 268
- del movimiento, II: 23
- de oscilación, II: 35
- parcial derivada, I: 376; II: 15
- soluciones admisibles o generalizadas de, II: 65, 69
- construcción de soluciones, II: 43
- paramétrica, II: 81
- del plano, I: 268
- de Poisson, II: 30
- de primer grado, I: 231, 320
- de Ricatti, I: 395
- secular, III: 106
- de segundo grado, I: 320
- solución algebraica de, I: 320
- de superficie, I: 266
- de tercer grado, I: 321
- Efimov, N. V., II: 127
- Egorov, D. F., III: 47
- Einstein, A., I: 301; III: 131, 200, 216, 223-227
- Eje, de grupo, III: 338-339
 - de rotación, III: 312
 - de simetría, III: 312
- Elemento(s), asociado, de un anillo, III: 401
- cero, III: 348
- conjugados, de un grupo, III: 349
- de un conjunto, III: 18
- de longitud, III: 205
- de una matriz, III: 58
- indescomponible, III: 401
- inverso, III: 99
- neutro, III: 348
- primo, III: 401
- de superficie, II: 120, 132
- unidad, III: 348-349
- Elipse, I: 240-241
 - imaginaria, I: 254
 - de inercia, I: 243-245
- Elipsoide, I: 269
 - imaginario, I: 269
- Envolvente de una familia de curvas, II: 135
- Equivalentes, caminos, III: 360
- Eratóstenes, I: 58; II: 241
- criba de, I: 58; II: 241
- Escalar, I: 262; II: 16
 - multiplicación, III: 80
- Escalar, producto, espacio n -dimensional, III: 271
- de vectores, I: 262; III: 78, 383
- Espacio, abstracto, III: 164
 - transformación de, I: 280-281; III: 162
 - base de, III: 71-72
 - configuración de, III: 213
 - de color, III: 185
 - curvatura de, III: 209-210
 - de curvatura cero, III: 211
 - curvatura constante, III: 211
 - dimensión de, III: 71, 259
 - de dimensión infinita, III: 71, 191, 196, 273-279, 298
- Euclidiano, III: 80
 - fásicos, I: 410; III: 187, 243
 - funcional, III: 192, 265, 299
 - de infinitas dimensiones, III: 299
 - generalizado, III: 190
 - de Hilbert, III: 196, 273-379
 - lineal complejo, III: 76
 - lineal general, III: 68
 - lineal de dimensión infinita, III: 298
 - lineal de dimensión finita, III: 386
 - métrico, III: 196, 262-263
 - multidimensionales, III: 168, 174
 - n -dimensional, III: 175

- real, III: 215
- de Riemann, III: 199-202
- rotación de, III: 318
- solución, III: 97
- topológico, III: 194, 262, 265
- topológico abstracto, III: 164
- unitario, III: 82
- vectorial lineal, III: 68-69
- Espacio-tiempo, III: 222
- Espectro continuo, III: 304-305
- Esperanza matemática, II: 287
- Estimación de un error, II: 360
- Estabilidad de los procesos de cálculo, II: 369
- Euclides, I: 22, 34, 40, 41, 46, n. 13 p. 46, 57, 58, 63, 77; II: 241, 244, 247; III: 123-130, 140, 141, n. 7 p. 144, n. 10 p. 151, 152, 154, 160, 164, 211, 215, n. 37 p. 215, 217, 220
- Euclides, postulado de las paralelas de, III: 124, 153
- Eudemo de Rodas, I: 39
- Eudoxio, I: n. 12 p. 46
- Euler, L., I: 75, 78, 98, 127, 201, 253, 259, 260, 317, 318, 402, 403, 407; II: 54, 78, 103, 105, 107, 114, 125, 134, 151, 154-157, 159, 161, 175, 176, 240, 242, 244, 246, 249, 260, 315, 343, 370; III: 243-247, 252, 256 281
- Euler, ángulos de, I: 269
- características de triangulación, III: 244
- ecuación diferencial de, II: 151, 154
- fórmula(s) de, II: 175; III: 244
- función(es) de, II: 260
- método de las líneas quebradas de, I: 403
- teorema de, II: 103, 105
- Euler-Poincaré, fórmula de, III: 252
- Excentricidad de la sección cónica, I: 251
- Existencia de una solución, I: 401
- Extremas, condiciones, II: 150
- Extremos, valores de funcionales, II: 149
- Extrínseca, propiedad de una superficie, II: 127
- F
- Faber, G., II: 323
- Factor irreducible, I: 334
- Factorización de un polinomio, I: 326
- Falsa, posición, I: 366, 368
- Familia de curvas, II: 134
- n -paramétrica, II: 134
- de superficies, II: 134
- Fedorov, E. S., I: 42, 82, 87; III: n. 23 p. 180, 311, 340, 342, n. 2 p. 343
- Fejér, L., II: 349, 352
- Fejér, suma de, II: 349
- Fermat, P., I: 226, 316, 317; II: 265; III: 402
- Ferrari, I: 63, 324
- Figuras, clasificación afín de, I: 283
- planas infinitas, III: 337
- propiedades afines de, I: 283
- Finikov, S. P., II: 143
- Finsler, R., I: 81; III: 214
- Flujo, II: 20
- laminar, II: 188
- rotacional, II: 26
- vector de, II: 20
- Foco, en ecuaciones diferenciales, I: 415; III: 258
- de una sección cónica, I: 247
- Fok, V. A., III: 226
- Fokker, II: 307
- Forma cuadrática, III: 109, 114
- canónica, III: 113
- equivalente, III: 111
- fundamental, III: 204
- inercia de una, III: 113
- primera fundamental, II: 131-134
- segunda fundamental, II: 133
- Forma lineal, III: 58
- Fórmula, de los incrementos finitos, I: 154
- fundamental del cálculo integral, I: 205
- Fourier, J. B. J., I: 358; II: 44, 54, 58, 315, 339, 341, 344, 347-349; III: 53, 196, 304
- Fourier, coeficientes de, II: 344
- integral de, III: 304
- series de, II: 341-349

- Fracciones, origen de, I: 43
 Frecuencia característica, *véase* Autofunción
 Frecuencias, espectro de, III: 293
 Frecuencia propia, III: 293
 fundamental, II: 51
 Frenet, J. F., II: 79
 Fridman, A. A., III: 227
 Frontera, de una cadena, III: 249
 de un conjunto, III: 194; *véase también* Punto frontera
 de un triángulo orientado, III: 247-248
 Fubini, G., II: 143
 Función(es), admisible, II: 152
 analítica, I: 219; II: 170
 completa, II: 224
 en intervalos, I: 220
 en un punto, II: 177
 representación de, II: 221
 aproximación de, II: 311
 aproximadora, II: 331
 armónica, II: 30
 máximo o mínimo absoluto de, II: 32
 característica, II: 47; III: 49
 de flujo, II: 187
 cociente de diferencia de, II: 60
 compleja, II: 167-231
 compuesta, II: 130
 constante a trozos, II: 68
 continua, I: 117
 a un lado, I: 119
 no regular, II: 139, 323
 continuamente diferenciable, II: 71; III: 55; *véase también* función continua
 de corriente, II: 185
 de cuadrado integrable, III: 299
 definición implícita de una, I: 181
 derivada de, I: 121
 derivable, I: 129; II: 195
 diferencial de, I: 150
 de Dirichlet, III: 46
 discontinua, I: 119
 dominio de, I: 103
 elemental, I: 139; II: 169
 entera, II: 216
 esférica, III: 281
 de Euler, II: 260
 exponencial, II: 174
 de matrices, III: 118
 fraccionaria, II: 218
 de una función, I: 137
 gráfica de, I: 103, 140-9
 de Green, III: 288-292
 impar, II: 237
 implícita, definición de, I: 181
 implícita, derivación de, I: 187
 integrable, I: 172, 196
 integral, II: 215
 inversa, I: 135
 de Lagrange, II: 70
 límite, III: 16
 lineal, I: 102; III: 55
 lineal homogénea, III: 56
 lisa, II: 68; III: 55, 98
 lisa a trozos, II: 68
 de una matriz, III: 117
 argumento, III: 117
 medible, III: 45
 multiforme, II: 181, 225
 no regular, II: 139
 ortogonales, III: 278
 ortogonales dos a dos, III: 279
 par, II: 237
 prolongación de, en un dominio complejo, II: 170
 regular, II: 322
 sistema ortogonal cerrado de, III: 287
 sistema ortogonal infinito de, III: 287
 de teoría de números, II: 260
 teoría trascendente de, II: 195
 trascendente, II: 217
 trigonométrica, II: 174
 de una variable compleja, II: 174
 de variable real, III: 15
 de varias variables, I: 179
 vectorial, II: 130
 Funcional, I: 155; III: 297

G

- Galerkin, B. G., II: 58, 59, 73, 358
 Galerkin, método de, II: 58
 Galileo, I: 65, 68, 70, 92, 93, 225, 301; II: 300, 301

- Galois, E., I: 82, 318, 330-8; III: 343-347
- Galois, grupo de, I: 355; III: 343-345
- teoría de, I: 332-6
- Gauss, K. F., I: 98, 318, 332, 336; II: 111-114, 128, 130, 134, 240, 244-330; III: n. 2 p. 126, 127, 131, 199, 209, 210, 216
- curvatura de, II: 108, 111; III: 209-210
- teorema de la divergencia de; véase Ostrogradski, fórmula de
- Gelfand, I. M., III: 358
- Gelfond, A. O., II: 240
- Generadores, de grupos, III: 363
- libres, III: 364
- de subgrupo, III: 330
- Generalizadas, soluciones de ecuaciones diferenciales parciales, II: 67, 69
- Generadores, de grupo, III: 363
- Generatrices rectilíneas, I: 274
- Género de una superficie, II: 230; III: 236
- Geodésica, II: 122; III: 143, 206
- Geodésica(o), curvatura, II: 119
- línea, II: 121
- local, II: 137
- superficie, III: 210
- segmento, II: 119
- Geometría, I: 38, 80
- abstracta, III: 215
- afín, III: 162
- algebraica, I: 311
- analítica, I: 68, 70, 225, 228, 311
- en el espacio, I: 266-7
- analítica de infinitas dimensiones, I: 311
- axiomas de la, III: 151
- conforme, II: 143
- diferencial, II: 77, 80-82, 143
- afín, II: 143
- clásica, II: 137
- proyectiva, II: 143
- euclidiana, III: 141
- global, II: 137
- intrínseca, II: 114-134; III: 201
- de Lobachevski, III: 132
- n-dimensional, III: 169
- no euclidiana, III: 123
- proyectiva, I: 293, 299; III: 159
- riemanniana, III: 199
- Geométrico(a), figura, I: 38
- interpretación, I: 398-400
- Gibbs, J. W., III: 174
- Gjunter, N. M., II: 72
- «Global», I: 420; II: 77, 136
- Goldbach, C., II: 249, 256, 259, 264
- Goldbach, conjetura de, II: 249, 256
- Gradiente de temperatura, II: 28
- Grado, de una representación mediante matrices, III: 367-392
- de simetría, III: 319
- Graduada, representación, III: 370
- Graeffe, I: n. 14 p. 369
- Graeffe, método de, I: 369
- Gráfica, de una función, I: 103, 104, 140-149
- de un polinomio, I: 364
- Grassmann, H., II: 129; III: 169
- Green, G., III: 288
- Green, función de, III: 288-292
- teorema de; véase también Ostrogradski, fórmula de
- Grupo(s),
- abeliano, III: 332
- abstracto, III: 373
- alternado, III: 327
- de automorfismos, III: 345
- axiomas de, III: 348
- de caminos, III: 361
- carácter de, III: 367
- clases disjuntas en, III: 349
- composición de, III: 347
- conmutativo, III: 332
- caracteres de, III: 261
- continuo, III: 356, 358
- cristalográfico, III: 334, 340, 342
- definición, general, III: 347
- derivado, III: 333
- de ecuaciones algebraicas, III: 344
- eje de, III: 338-339
- elemento conjugado en, III: 349
- factor, III: 351
- de Fedorov, III: 334, 340, 342
- finito, III: 373
- finito simple, III: 358
- fundamental, III: 359

de Galois, I: 355; III: 343-344
 generadores de, III: 363
 infinito, III: 373
 isomorfo, III: 349
 libre, III: 364
 de Lie, III: 356, 358
 local, III: 396
 de movimientos, III: 338
 multiplicación de, III: 347
 de nudos, III: 365, 366
 operación de, III: 348
 orden de, III: 325
 de permutaciones, III: 327
 plano de Fedorov, III: 334, 340, 341, 342
 relaciones de definición de, III: 363
 representación de, III: 367
 representación graduada en, III: 370
 representación por matrices, III: 367
 de rotaciones, III: 326
 simétrico, de permutaciones, III: 327, 347
 suma de, III: 348
 simple, III: 352
 topológico, III: 358
 de transformaciones, III: 321-334, 347
 en geometría diferencial, II: 143
 Gyaseddin Jamschid, I: n. 18 p. 60, 62

H

Hadamard, J., II: 246
 Hamilton, W. R., II: 70-72, 130, 160
 Hamilton, principio de; véase Principio de Ostrogradski-Hamilton
 Hardy, G. H., II: 240
 Hermitiano, espacio vectorial; véase Espacio unitario
 Hertz, H., I: 21
 Hilbert, D., I: 84, 99, 337, 338
 espacio de, III: 196, 273-279
 Hincin, A. Ja., I: 425; II: 289
 Hiparco, I: 57

Hipérbola, I: 247
 rectangular, I: 253
 Hiperboloide, I: 269, 273
 de revolución, I: 273
 Hipercomplejos, números, III: 374-385
 sistemas, III: 310, 385
 Hiperplano, III: 171
 tangente, III: 184
 Hipócrates, I: 56
 Holónomo, sistema, II: 160
 Homeomorfas, figuras, III: 235
 Homología, III: 250
 Homológica, independencia de los ciclos, III: 250
 teoría de la dimensión, n. 11 p. 260
 Homólogo a cero, III: 250
 Homomorfa, aplicación, III: 353
 imagen, III: 256
 Homomorfismo, III: 353
 Homotópica, topología, III: 261
 Hooke, R., III: 259
 Hopf, H., III: 256, 260, 261
 Hopf, teorema de, III: 256

I

Ideal, III: 399
 bilátero, III: 393
 cero, III: 393
 de elementos de un álgebra III: 392
 a derechas, III: 392
 a izquierdas, III: 393
 minimal, III: 393
 principal, III: 399
 indescomponible, III: 403
 Ideales, líneas, I: 297
 puntos, I: 298
 Imagen especular, III: 311
 Inconmensurables, magnitudes, I: 44
 Incremento, I: 118, 150
 Independencia, III: 57
 lineal, III: 69
 Independientes, sucesos, II: 275
 Índice, de un punto singular, III: 256

de un punto singular aislado, III: 256
 Infinitamente grande, I: 112
 Infinitésimo, I: 110
 de orden superior, I: 151
 Inflexión, punto de, I: 148
 Inicial, condición, II: 34
 estado, I: 402; II: 29
 valor de un problema, I: 402; II: 29
 Integración, de caminos, II: 210
 por partes, I: 178
 Integral, I: 163
 de acción, II: 161
 de Cauchy, II: 211
 a lo largo de una curva, II: 208
 curvilínea, I: 201-3
 definida, I: 165
 doble, I: 200
 indefinida, I: 173; II: 210
 de Lebesgue, III: 47-53
 de línea, I: 201; II: 207
 sobre un contorno cerrado, II: 211
 múltiple, I: 197
 de Riemann, I: 165; III: 47
 de superficie, I: 201
 Interpolación, lineal, I: 366, 368
 polinomios de, II: 316, 335
 Intersección, de conjuntos, III: 19-20
 de sucesos, II: 274
 Intervalo, III: 33
 abierto, III: 33
 adyacente, III: 38
 básico, II: 259
 cerrado, III: 33
 encajado, III: 29
 impropio, III: 33
 Invariante, I: 288
 topológico, III: 362
 Inverso, de un cuaternio, III: 381
 de una matriz, III: 93
 Irreducibilidad de un polinomio, I: 334
 Irrotacional, movimiento en un fluido, II: 187
 Isomorfa(s), aplicación, III: 350
 Isomorfismo, III: 349
 local, III: 396
 Iteración, método de, II: 362

J

Jackson, D., II: 314, 340
 Jordan, C., III: 108, 121
 Jukovski, N. E., I: 22

K

Keldys, L. V., III: 41
 Keldysch, M. V., II: 340; III: 305
 Kepler, J., I: 70, 71, 91, 96, 226, 242, 243, 375
 Killing, III: 358, 398
 Kirchhoff, G. R., II: 389
 Kiselev, A. P., I: 40
 Klein, F., I: 99, 338; III: 144, 150-154, 163, 200, 214, 225, 238, 257, 258, 358
 Klein, botella de, III: 256-257
 Kolmogorov, A. N., II: 307, 314, 315, 348; III: n. 1 p. 25, 41, 261
 Kolosov, G. V., II: 195
 Korkin, A. N., II: 314
 Krasnoselski, M. A., III: 265
 Krylov, A. N., I: 372, 405
 Krylov, N. M., I: 424, 425
 Kuiper, N. H., II: 142; III: n. 8 p. 144
 Kummer, E. E., II: 240; III: 402
 Kurnakov, N. S., III: 174
 Kurosch, A. G., I: 320; III: 373

L

Lagrange, J. L., I: 74, 78, 154, 193, 194, 195, 196, 260, 261, 318, 328-32, 334, 335, 337; II: 70, 240, 320, 329; III: 16, 127, 169, 223, 343, 352
 Lagrange, fórmula de, I: 154
 fórmula de interpolación de, II: 320
 función de, II: 70
 multiplicadores de, I: 78, 193
 resolventes de, I: 330
 teorema de, III: 352
 Laguerre, E., I: 260
 Lambert, J. A., III: 125, 126

- Laminar, flujo, II: 188
 movimiento, II: 183
 Laplace, P. S., II: 30-32, 40, 43, 52, 54, 62, 68, 194, 284, 303, 359, 382, 383
 ecuación de, II: 30
 Laplaciana, II: 30
 Lappo-Danilevski, I. A., III: 121
 Laptev, G. F., II: 143
 Lavrentiev, M. A., I: 404; II: 207, 314, 340; III: 41
 Lebesgue, H. I.: 83, n. 25 p. 169, 197; II: 72, 314, 348, 353; III: 17 41, 47-53, 279, 287
 suma integral de, III: 52
 Lefschetz, S., III: 260
 Legendre, A. M., II: 244, 330; III: 125, 126, 282
 polinomios de, II: 330; III: 282
 Leibnitz, G. W., I: 68, 71, 96, 97, 127, 167, 169, 170, 171, n. 27 p. 171, 173, 316; II: 78, 135; III: 131
 Leverrier, U. J. J., I: 21, 375
 Ley(es) de inercia, III: 113
 de los grandes números, II: 280, 289
 de probabilidad, II: 269
 triangular, III: 264
 L'Hôpital, II: 78
 Lie, S., I: 320; III: 261, 356-358, 389, 395-398
 Lie, álgebra de, III: 395
 grupo de, III: 356-358
 representación continua de, III: 358
 Límite, I: 72, 108-17
 de una sucesión de matrices, III: 117
 uniforme, II: 69
 Línea, ecuaciones de, I: 268
 infinitamente distante, III: 239
 en el infinito, III: 239
 impropia, III: 239
 integral, I: 399
 Linnik, Ju. V., II: 247; III: 385
 Lisma, II: 142
 Liouville, J., I: 333, 395
 Littlewood, J. E., II: 240
 Liapunov, A. M., I: 74, 84, 99, 416-9, 424; II: 58, 290; III: 41
 estable en el sentido de, I: 417
 no estable en el sentido de, I: 417
 Ljusternik, L. A., II: 138; III: 260, 262, 265
 Lobachevski, N. I., I: 22, 40, 80, 81, 99, 108, 310, 318, 369, 371, 372; III: 123, 127-159, 164, n. 28 p. 193, 211, 215, 216, 217, 220, 221, 222, 227
 geometría de, III: 132
 curva equidistante en la, III: 135
 método de, I: 369
 postulado de las paralelas de, III: 154
 «Locales», II: 77, 136
 Lógica matemática, I: 86
 Lomonosov, M. V., III: 185
 Lorentz, H. A., I: 276, 301, 304-10; III: 225
 transformación de, I: 301-3
 fórmulas de transformación de, I: 304
 Luzin, N. N., I: 83; II: 315; III: 41, 47

M

- Mach, E., III: 222
 Malsev, A. I., III: 358
 Mandelstam, L. I., I: 424; III: 254
 Máquina integradora, II: 381
 clasificadora, II: 378
 digital, II: 378
 Máquinas analógicas o de acción continua, II: 380
 Mariotte, E., I: 105, 293; II: 29; III: 174
 Markov, A. A., I: 99, 425; II: 289, 290, 306, 314, 336
 Markov, V. A., II: 314
 Markov, proceso de, II: 306
 Markusevic, A. L., III: n. 9 p. 259
 Matemática moderna, I: 87
 Matricial, álgebra, III: 389
 cálculo, III: 117
 elemento, III: 58
 notación, III: 92
 operación, III: 59

- representación, de álgebras asociativas, de grupos, III: 367
- Matriz (matrices), III: 58
 - antisimétrica, III: 397
 - canónica de Jordan, III: 108
 - de coeficientes, III: 62
 - determinante de, III: 90
 - columna de, III: 59
 - cuadrada, III: 59
 - potencias de, III: 117
 - determinante de, III: 83
 - diagonal, III: 108
 - ecuación característica de, III: 106
 - de igual estructura, III: 62
 - fila de, III: 59
 - función, III: 117
 - exponencial, III: 119
 - multiplicación de, III: 63
 - no degenerada, III: 94
 - no nula, III: 401
 - no singular, III: 94
 - nula, III: 63, 64
 - números característicos de, III: 106
 - ortogonal, III: 397
 - polinomio característico de, III: 106
 - rango de, III: 95
 - semejante, III: 104
 - de transformación de coordenadas, III: 103
 - traspuesta, III: 64
 - unidad, III: 93, 391
- Máximo, I: 140
- Máximo, valor de una función, I: 141
- Maxwell, J. C., I: 21; III: 187
- Medible, conjunto, III: 43
 - función, III: 45
- Medida de un conjunto, III: 41, 43
 - de un conjunto abierto, III: 42
- Medida exterior, III: 43
- Mendelciev, D. I., II: 336
- Menshov, D. E., II: 315
- Mergeljan, S. N., II: 340
- Meromorfa, función, II: 219
- Método, de aproximaciones sucesivas, I: 405; II: 362
 - combinatorio en topología, III: 243-244
- de coordenadas, I: 229
- del «corto» y «largo», I: 366
- de las cuerdas, I: 366, 368
- de diferencias finitas o de redes, II: 58, 60
- geométrico, III: 197
- de las líneas quebradas (véase de las poligonales)
- de Newton, I: 368
- numéricos, II: 355
- de las poligonales, I: 403; II: 165, 166
- de las potenciales, II: 54
- de redes, II: 58, 60
- de sustitución, I: 175
- de las tangentes, I: 366
- Métrica, III: 201
 - intrínseca, III: 196
- Meusnier, J. B. M., II: 104, 105, 107, 108, 114, 134
 - teorema de, II: 104
- Michelson, A. A., I: 301, 302
- Miguel Angel, I: 91
- Minding, F., II: 79, 125, 137; III: 143
- Minimal(es), ideal, III: 393
 - principios, II: 130
 - problema, II: 147
 - superficie, II: 111
- Minimizadora, sucesión, II: 165
- Mínimo, I: 140
 - valor de una función, 141
 - problema del, para una integral múltiple, II: 161
- Minkowski, H., II: 139, 240; III: 223
- Möbius, A. F., I: 298; III: 237-240
 - banda de, III: 236-237
 - cinta de, III: 237
 - red de, I: 298, 299
 - superficie de, III: 236-238
- Modelo circular del plano proyectivo, III: 240-241
- Modular, retículo, III: 406
- Módulo, de un número complejo, I: 340
 - de un polinomio, I: 345
 - superficie de, I: 345, 351
- Moivre, de, II: 284
- Molin, F. E., III: 394, 395, 404

Monge, G., II: 137
 Monotonía, ley de, III: 27
 Moreland, II: 376
 Morozov, V. V., III: 358
 Morse, M., II: 138; III: 265
 Movimiento, ecuaciones de, II: 23
 helicoidal, III: 318
 impropio, III: 318, 342
 permanente, II: 183
 no perturbado, I: 418
 perturbado, I: 418
 plano-paralelo, II: 183
 de primera especie, III: 318
 propio, III: 318
 de segunda especie, III: 319
 Multiplicación, tabla de, III: 350
 Multiplicador, II: 378
 Multiplicativa, propiedad, II: 238
 Multiplicidad de una raíz, I: 353
 Mushelishvili, N. I., II: 195

N

Naimark, M. A., III: 358
n-dimensional, ángulo, III: 170
 conjunto, III: n. 11 p. 260
 cubo, III: 176
 cuerpo, III: 181
 curva, III: 184
 espacio, III: 268-269
 espacio euclidiano, III: 77, 80
 geometría, III: 170
 plano, III: 170
 punto, III: 66
 poliedro, III: 175-181
 segmento, III: 170
 simplex, III: 178, 247
 topología, III: 259
 triangulación, III: 247
 vector, III: 66, 270
 Nash, J., II: 142
 Nasireddin Tusi, I: 61; III: 125
 Nemyski, V. V., I: 425; III: 255, 265
 Neper, J., I: 64, n. 21 p. 64, 68
 Newton, I., I: 54, 61, 62, 64, 68, 71, 73, 74, 75, 78, 96, 97, 116, 167-73, 178, 228, 237, 243, 260, 316, 317, 318, 358, 368, 369, 375,

378; II: 23, 25, 29, 78, 176, 300, 313, 320; III: 131
 método de, I: 368
 potencial de, II: 55
 teoría de los diámetros de, I: 237
 Newton-Leibniz, fórmula de, I: 169
 No enumerabilidad del continuo, III: 32
 No orientabilidad, III: 246
 Nodo(s), I: 414; III: 253
 equivalentes, III: 365
 no equivalentes, III: 365
 de una oscilación, II: 51
 Normal, II: 99
 principal, II: 95
 Novikov, P. S., III: 41, 367
 Núcleo, de una ecuación integral, III: 294
 simétrico, III: 295
 Nudos, III: 365
 equivalentes, III: 365
 no equivalentes, III: 365
 Numerabilidad, III: 27
 Número(s), I: 27
 algebraico, II: 240
 analítico, II: 340
 de Betti, bidimensional, III: 251
 característico, III: 106
 cardinal del continuo, III: 25
 compuesto, II: 242
 de dirección, I: 407
 entero, I: 24; II: 235
 negativos, I: 55
 primo, II: 235, 237
 real, axiomas de continuidad de, I: 347; III: 29, 153
 teoría analítica de, II: 340
 elemental de, II: 339
 geométrica de, II: 339
 de vueltas del campo, III: 256

O

Odner, V. T., II: 377
 Omar Khayyam, I: 61, 62
 Onda esférica divergente, II: 43
 Operación(es), sobre conjuntos, III: 19

- sobre determinantes, III: 86
- m-aria, III: 407
- no conmutativa, III: 321
- sobre vectores, III: 65
- Operador(es), III: 299
 - autoadjunto, III: 297, 302, 303
 - lineal, III: 303
 - lineal, III: 296, 299
- Orden, de conexión, III: 236, 251
 - de un conjunto de puntos, III: 33
 - de contacto, I: 404
 - de ecuaciones diferenciales, I: 384
 - de grupo, III: 325
 - de grupo simétrico, III: 327
 - de un infinitésimo, I: 151
 - de una matriz cuadrada, III: 59
 - de simetría, III: 312
 - de tangencia, I: 404
 - de transformación, III: 331
- Ordenador, I: 260
- Orientabilidad, III: 245
- Origen, I: 227
- Ortogonal(es), complementos, III: 82
 - funciones, III: 278
 - invariante, I: 289
 - matriz, III: 397
 - proyección, III: 82
 - sistema, II: 351; III: 279, 280
 - transformación, I: 287; III: 101
 - de una forma cuadrática, III: 114
- Ortogonalidad, II: 47
 - de vectores, III: 80, 270
 - propiedad de, II: 345
- Ortonormal, base, III: 80
 - sistema de funciones, III: 281
- Oscilación(es), armónica, III: 292
 - característica, II: 47; III: 293
 - ecuaciones de, II: 35
 - forzada, I: 394
 - libre, III: 292
 - modo de, II: 51
- Ostrogradski, M. V., I: 74, 99, n. 26 p. 169, 179, 203, 205; II: 58, 159
 - fórmula de, I: 203; II: 21, 163
 - Hamilton, principio de, II: 159
- Papaleksi, N. D., I: 424
- Parábola, I: 248
- Parábola de orden n , I: 364
- Paraboloide, I: 273
 - elíptico, I: 269
 - hiperbólico, I: 267, 270
- Paralelas, líneas, III: 133-135
- Paralelas, postulado de, III: 124
- Paralelismo, III: 133
- Parámetro, I: 311
- Paridad, III: 328
- Parte principal, II: 220
- Parseval, M., III: 53
- Parseval, igualdad de, III: 53
- Pascal, B., II: 376
- Perforadora reproductora, II: 378
- Permutación, impar, III: 327
- Perspectiva plana, teorema fundamental de; véase Teorema de unicidad de la teoría de transformaciones proyectivas
- Peterson, K. M., II: 128, 134
 - teorema de, II: 79, 128
- Petrovski, I. G., I: 312; II: 307
- Picard, E., II: 217
- Picard, método de; véase Método de aproximaciones sucesivas
- Pitágoras, I: 40, 45, 51, 69, 229, 240; III: n. 6 p. 140, 202
- Planck, M. K. E. L., II: 307
- Plano, completado, I: 297
 - cubierto densamente, I: 299
 - ecuaciones de, I: 268
 - de las fases, I: 409
 - n -dimensional, III: 169
 - normal, II: 94
 - osculador, II: 94
 - proyectivo, en geometría, I: 297
 - en topología, III: 239-240
 - de simetría, III: 311
 - tangente, II: 98-101
- Pogorelov, A. V., II: 141
- Poincaré, H., I: 74, 84, 99, 301, 416, 424; III: 131, 151, n. 11 p. 151, 154, 222, 243, 252, 254, 258, 260
- Poisson, S. D., II: 30-32, 43, 52, 54, 58
- Poliedros, II: 140

- en un espacio n -dimensional, III: 175
 Polinomio(s), característico, de una matriz, III: 106
 gráfica de, I: 364
 de interpolación, II: 316, 335
 irreducibilidad de, I: 334
 de Legendre, II: 330; III: 282
 de una matriz, III: 117
 simétrico, I: 328; III: 320
 simétrico elemental, I: 328
 trigonométrico, II: 314
 Polo II: 219
 Poncelet, J. V., III: 160
 Pontriagin, L. S., I: 424; III: 260.
 n. 12 p. 260, 261, 311, 359
 ley de dualidad de, III: 261
 Popov, A. S., I: 21
 Posición falsa, método de, I: 366, 368
 Postnikov, M. M., III: 261
 Potencial, de capa simple, II: 56
 de fuerzas, II: 55
 de velocidad, II: 27, 187
 de un fluido, II: 26
 Primitiva, función, I: 168
 Primo, gemelo, II: 242
 Principio(s), del argumento, I: 360
 de Arquímedes, III: 31-32
 de continuidad de Cantor, III: 30
 de Dedekind, III: 29
 Probabilidad, II: 269, 273
 adición de, II: 273
 densidad de, II: 306
 distribución de, II: 286
 Proceso aleatorio, II: 299, 306
 Proclo, III: 125
 Producto, III: 26
 de caminos, III: 361
 de ciclos, III: 328
 derivada de, I: 133
 libre, III: 364
 de matrices, III: 60
 no conmutativa, III: 60
 de transformaciones, III: 321
 Prolongación analítica de una función en un dominio complejo, II: 170-171, 222-225
 Propagación de ondas, II: 40
 Propiedad intrínseca, II: 127
 Proporcionalidad, III: 56
 Proximidad, III: 359
 Proyección perspectiva, I: 293-5
 Proyectiva(s), aplicaciones: teorema fundamental, I: 298
 transformación, I: 298; II: 143
 Punto(s)
 adherente, III: 262-264
 aislado, III: 35
 conjunto de, III: 32
 conjunto lineal de, III: 32
 coordenadas de, I: 227
 de discontinuidad, I: 411
 de ensilladura, I: 413
 de estancamiento, II: 193
 fijo, III: 257
 frontera, III: 35
 impropio, III: 239
 interior de un conjunto, III: 35
 de interpolación, II: 320
 de intersección, I: 230
 de inflexión, I: 148
 límite, III: 35
 ordinario, I: 410
 plano, II: 103
 simétrico, III: 311
 singular, III: 253, 256
 aislado, I: 410
 de tangencia, II: 99

R

- Racionales, números, III: 26
 Radical, III: 394
 de un anillo, III: 404
 Radio de convergencia, I: 220; II: 173; III: 160
 Radio vector, II: 130
 Rafael, I: 91
 Raíces, cálculo de, I: 364
 múltiples, I: 353
 multiplicidad de, I: 353
 reales, I: 327
 Rama uniforme, II: 182, 227
 Ramanujan, S., II: 240
 Ramificación, punto de, de función compleja, II: 182, 226
 Rango, de álgebra, III: 387
 de una matriz, III: 95

- de números hipercomplejos, III: 376
 - de un sistema de vectores, II: 75
 - Rasevski, P. K., III: 215
 - Real, espacio, III: 215
 - número, I: 47; III: 25, 28
 - parte de un cuaternio, III: 381, 382-383
 - punto, I: 298
 - raíz, I: 327
 - recta, I: 298
 - Reciprocidad, ley de, III: 289
 - Recta, línea, II: 123
 - Recta numérica, III: 16, 29
 - Redes de curvas, II: 79
 - Reducción, a una forma canónica, I: 253
 - de una forma cuadrática, III: 111
 - representación, III: 370
 - Reducible, ecuación, I: 292
 - Regular, representación, III: 371
 - sistema de puntos espaciales, III: 342
 - superficies, II: 97
 - Representación completamente reducible, III: 370
 - Representación fiel, III: 391
 - Representación(es), de álgebras asociativas, III: 391
 - equivalentes, III: 367
 - de grado n , III: 368
 - de grupo, III: 367
 - irreducible, III: 370
 - isomorfa, III: 391
 - Reproductora, II: 378
 - Resolubilidad, de grupo, III: 334
 - de problemas con regla y compás, I: 335
 - mediante radicales, III: 346
 - Resoluble, grupo, III: 334
 - mediante radicales, I: 337
 - Reticulo, III: 404
 - modular, III: 406
 - «Resto», término, I: 159
 - Riccati, J. F., I: 395
 - ecuación de, I: 395
 - Riemann, B., I: 80, 81, 99, 121, 312; II: 180, 183, 187, 194, 198, 201-207, 218, 227-230, 240, 246; III: 16, 45, 47, 49, n. 7 p. 144, 197-214, 216, 221, 222, 226, 258, 357
 - espacio de, III: 199
 - geometría de, III: 199
 - hipótesis de, II: 246
 - integral de, I: 165; III: 47
 - métrica de, III: 205
 - superficie de, II: 201-207; III: 258
 - Riesz, F., I: 84
 - Ritz, W., II: 73-74, 359
 - Rolle, M., I: 354, 356
 - teorema de, I: 354
 - Rotación, del espacio, III: 318
 - de una transformación, II: 197
 - Rotaciones, grupo de, III: 326
 - Rotacional, flujo, II: 26
- S
- Saccheri, G. G., III: 126, n. 2 p. 126, 128, 131
 - Safarevich, I. R., I: 338; III: 346
 - Schauder, J., III: 265
 - Schlaefli, L., III: 176
 - Schrödinger, E., II: 73
 - Schwartz, desigualdad de; véase Desigualdad de Cauchy-Bunjakovski
 - Schweikart, III: 127, 131
 - Sección normal, II: 102
 - Segmento parabólico, I: 404
 - Selberg, A., II: 246
 - Semiintervalo, III: 33, 39
 - Semiinvariante, I: 290
 - Semirretículo, III: 405
 - inferior, III: 405
 - superior, III: 405
 - Separación de variables, II: 44
 - Series, I: 206
 - absolutamente convergente, I: 211
 - armónicas, I: 210
 - divergentes, I: 207
 - no uniformemente convergentes, I: 213
 - de potencias, I: 217; II: 169
 - sumables, I: 207, n. 37 p. 207
 - de Taylor, I: 162
 - trigonométricas, III: 283

- uniformemente convergentes, I: 212
- Serre, J. P., III: 261
- Serret, J. A., I: 318
- Seudoesfera, III: 142
- Símbolos, I: 27
- Simetría, III: 311, 317
- bilateral, III: 312
- central, III: 312
- centro de, III: 312
- circular dirigida, III: 336
- eje de, III: 312
- especular, III: 311
- de orden n , III: 312
- respecto a ejes, III: 342
- Simplex, III: 178, n. 22 p. 178
- orientado, III: 250
- Simpson, T., II: 324-330, 360, 368, 369
- fórmula de, II: 325
- Singular, punto, I: 410; III: 256
- Sistema físico, oscilaciones características, I: 391; III: 293
- Sistema, algebraico, III: 309, 406-408
- binario, III: 404
- de ecuaciones lineales, III: 83
- homogéneo de ecuaciones lineales, III: 96; *véase también*
- Ecuaciones de primer grado de intervalos, III: 29
- normal, II: 351
- ortogonal de funciones, III: 282, 285
- normal, II: 351
- posicional, I: 29
- Sitnikov, K. A., n. 12 p. 260, 261
- Smidt, O. Ju., III: 311, 373
- Smirnov, Ju. M., III: 266
- Smirnof, V. I., I: n. 33 p. 194
- Snirelman, L. G., II: 138; III: 260
- Sobolev, S. L., II: 69, 73
- Solución(es), admisible, II: 68, 69
- algebraica, I: 320
- de ecuaciones diferenciales, I: 384
- generalizada, II: 68, 69
- unicidad de, I: 401; II: 29
- Stepanov, V. V., I: 425; II: 54, 302; III: 255
- Stevin, S., I: 64
- Sturm, S., I: 318, 358
- teorema de, 357
- Subálgebra, III: 393
- Subanillo, III: 399
- Subespacio, III: 74
- Subespacio generado por un sistema de vectores, III: 74-75
- Subgrupo(s), III: 330-334
- generadores de, III: 330
- invariante, III: 352
- normal, III: 351
- normales no triviales, III: 358
- simple, III: 358
- Subnikov, A. V., III: 343
- Sucesión, convergente en casi todo punto, III: 46
- fundamental, III: 30
- minimizante, II: 16
- de números reales, III: 30-31
- Suceso(s), aleatorio, II: 271, 296, 303
- cierto, II: 269
- estocástico, II: 271
- imposible, II: 269
- independiente, II: 275
- intersección de, II: 274
- mutuamente excluyentes, II: 272
- necesario, II: 269
- unión, II: 272
- Suma, de conjuntos, III: 19
- derivada de una, I: 132
- de Fejér, II: 349
- parcial, I: 207
- de series, I: 207
- trigonométrica, II: 247, 248
- de vectores, I: 262
- Sup., I: 213, n. 39 p. 213
- Superficie(s),
- analítica, II: 125
- cerrada(s), III: 237, 245, 246
- clase de, II: 139
- convexa, II: 140-141
- orientable, III: 250-251
- con curvatura variable de Gauss, II: 121
- conexión de, III: 236
- congruente, II: 128
- de curvatura media nula, II: 111
- de dos caras, III: 236
- de una sola cara, III: 236
- deformabilidad de, II: 142

- deformación de, II: 115, 125, 142
 - desarrollable, II: 120
 - de discontinuidad, II: 68
 - ecuaciones de, I: 266
 - familia de, II: 134
 - género de, II: 230; III: 236
 - geodésica, III: 210
 - lisa, II: 99
 - mínima, II: 111
 - de Möbius, III: 236-237
 - no cerrada convexa, II: 142
 - no lisa, II: 140
 - no regular, II: 137
 - normal, III: 236
 - regular, II: 97
 - de revolución de menor área, II: 147
 - teoría de, II: 96
 - Superposición, método de, II: 44
 - Suslin, M. Ja., III: 41
 - Sustracción, III: 26
 - Sylvester, J. J., I: 358; II: 245
- T
- Tablas, II: 273
 - de dos entradas, II: 374
 - de una entrada, II: 374
 - Tabulador, II: 378
 - Tales, I: 40
 - Tangente, I: 122; II: 86, 87
 - Tangentes, campo de, I: 399
 - Tartaglia, N., I: 63, 318, 321, 322
 - Taurinus, III: 127, 131
 - Taylor, B., II: 320, 321
 - fórmula de, I: 157, 191
 - serie de, I: 162
 - Temperatura, gradiente de, II: 28
 - Tensor, II: 16
 - Teorema,
 - de Abel, I: 332; II: 171
 - de Bolzano-Wierstrass, I: 347; III: 36
 - de Bonnet, II: 128
 - de Budan, I: 357
 - de Cauchy, II: 210
 - central del límite, II: 290
 - de Codazzi, II: 80, 128
 - de contracción, II: 364
 - de dualidad, III: 260
 - de Euler, II: 103, 105
 - fundamental, del álgebra, I: 338-352; II: 169
 - de la perspectiva plana, I: 295-297
 - de Gauss; véase Ostogradski, fórmula de
 - de Green, II: 163
 - de Hopf, III: 256
 - del límite central, II: 290
 - de Meusnier, II: 104
 - de Peterson, II: 79, 128
 - de Picard, II: 217
 - del punto fijo, de Brouwer, III: 198, 258
 - de Rolle, I: 354
 - de Sturm, I: 357
 - de Sukovski, II: 190, 193
 - de unicidad, para funciones analíticas, II: 222
 - de la teoría de transformaciones proyectivas, I: 294
 - del valor medio, I: 154
 - de Wedderburn, III: 395
 - de Weirstrass, II: 337
 - de Wilson, II: 265
 - Teoría, algebraica de números, II: 240
 - analítica de números, II: 240
 - cualitativa de ecuaciones diferenciales, I: 415; III: 253-254
 - de curvas, II: 80
 - elemental de números, II: 240
 - de funciones de variable real, III: 16-17
 - de Galois, I: 332
 - geométrica, de funciones de números, II: 240
 - de Newton, de diámetros, I: 237
 - de números, II: 235
 - de operadores, III: 297
 - Tihonov, A. N., III: 266
 - Tolomeo, C., I: 58, 60, 63
 - Topológico(a), aplicación, III: 263
 - espacio, III: 193, 262, 265
 - estructura, III: 261
 - grupo, III: 358
 - imagen, III: n. 10 p. 259
 - invariante, III: 247, 259, 362
 - producto, III: 242

- propiedad, III: 232
 - transformación, III: 232
 - Topología, III: 231
 - ciclos en, III: 249
 - combinatoria, III: 260
 - método de, III: 243-244
 - conjuntista, III: 259
 - n -dimensional, III: 259
 - orientaciones compatibles en, III: 245
 - de superficies cerradas, III: 258
 - Toro, III: 241
 - de una cara, III: 238
 - Torricelli, I: 97
 - Torsión, II: 95; III: 250
 - Tractriz, III: 142
 - Trascendente, función, II: 217
 - número, II: 240
 - Transformación(es), afín, I: 280; III: 162
 - bicontinua, III: 362
 - conforme, II: 197, 200
 - conjugada, III: 323, 324, 332, 333
 - de coordenadas, I: 255; III: 102
 - cuasi-conformes, II: 204
 - degenerada, III: 101
 - general afín, I: 280
 - grupo de, III: 351
 - de grupo, III: 163
 - grupos continuos de, III: 336
 - homogénea, III: 99
 - identidad, III: 330
 - inversa, III: 99
 - lineal, III: 98
 - de Lorentz, I: 301, 303
 - matriz de los coeficientes de, III: 100
 - no degenerada, III: 101
 - orden de, III: 331
 - ortogonal, I: 287; III: 101
 - periódica, III: 331
 - producto de, III: 321
 - proyectiva, I: 298; II: 143; III: 158
 - simétrica, III: 314
 - singular, III: 101
 - no singular, III: 101
 - topológica, III: 232
 - unitaria, III: 358
 - uno a uno, III: 315
 - Transposición, III: 328
 - Traslación, III: 318
 - Traspuesta, III: 64
 - matriz, III: 102
 - Traza, III: 372
 - Treffitz, J., II: 73
 - Triangulación, III: 244
 - tridimensional, III: 247
 - Triangular, desigualdad, III: 196
 - ley, III: 264
 - Tschirnhausen, E. W., I: 318, 326
- U
- Ulug Begh, I: 62
 - Unicidad, de descomposición, III: 401
 - propiedades de, II: 222
 - de solución, I: 401; II: 29
 - Unicidad, teorema de, para funciones analíticas, II: 222
 - de la teoría de transformaciones proyectivas, I: 295
 - Unidad, III: 401
 - Uryson, P. S., III: 259, 260, n. 11 p. 260, 263, 266
- V
- Vagner, V. V., III: 215
 - Vallée-Poussin, C. J., de la, II: 314
 - Van der Waerden, B. L., I: 320
 - Variable, I: 65
 - aleatoria, II: 286
 - cambio de, I: 175
 - continua, I: 54
 - dependiente, I: 103
 - independiente, I: 103
 - separación de, II: 44
 - Variable, magnitud, I: 65, 110
 - Variación, II: 154
 - Variacional, problema, II: 157
 - Varianza, II: 287
 - Varichak, III: 151
 - Variedades, III: 241-243
 - aplicación continua de, III: 257

- conexión de, III: 251
 - irreducible, III: 403
 - lisas, III: 255
 - multi-dimensionales, III: 259
 - Vavilov, S. I., I: n. 1 p. 92
 - Veblen, O., III: 260
 - Vector(es), I: 261; III: 65-66
 - combinación lineal de, III: 70
 - de un espacio infinito-dimensio-
nal, III: 275
 - flujo, II: 20
 - inverso de un, I: 262
 - longitud de, III: 80
 - n -dimensional, III: 270
 - dos a dos, III: 272
 - velocidad, II: 184
 - Vectorial(es), álgebra, I: 261; III: 381
 - campo, II: 15; III: 252
 - continuo, III: 252
 - coordenadas, I: 263; III: 73
 - función, II: 130
 - operaciones, III: 65
 - parte de un cuaternio, III: 380, 382
 - suma, I: 262
 - de dos subespacios, III: 75
 - Velocidad potencial, II: 27, 187
 - Vieta, F., I: 61, 63, 317, 319, 326, 328; III: 344
 - Vinogradov, I. M., I: 87; II: 240, 241, 246, 256-264, 316; III: 402
 - Vonoro, G. F., II: 241
- W
- Wachter, III: 127
 - Wallace, A. R., III: 131
 - Wallis, J., III: 125
 - Wedderburn, III: 394, 395, 404
 - teorema de, III: 395
 - Weierstrass, K. T., I: n. 14 p. 46, 83, 99, 129; II: 314, 337; III: 30, 31, 36
 - principio de, III: 30
 - teorema de, II: 337; III: 36
 - Weyl, H., III: 398
 - Wilczynski, E. J., II: 143
 - Wilson, teorema de, II: 265
- Z
- Zenón, I: 46
 - Zolotarev, E. I., I: 318; II: 240, 314; III: 402
 - Zukovski, N. E., I: 344; II: 190, 191, 193
 - teorema de, II: 190, 193